Excursions in Computing Science:
Week 8c. Symmetry: Simplifying Matrices.
Part IV Abstract symmetries and lots of physics.

T. H. Merrett[*]

McGill University, Montreal, Canada

October 6, 2009

I. Prefatory Notes

1. Which operations preserve the triangle?

2. How did we get the last two reflections?

3. We can abstract all this.

4. Other properties of the operators.

5. Operating on the group operators.

6. ghg^{-1}.

7. The rotation subgroup, $\{(), (123), (132)\}$.

8. Simplifying matrices.

9. Other representations of the group.

10. Simplifying matrices.

11. Regular representation.

12. Molecules

13. Modes of motion of a molecule.

14. Greenhouse gases.

15. The water vapour molecule.

16. Tetrahedron.

17. Hexa/Octahedra.

18. Dodeca/Icosahedra.

19. Infinite groups

20. 1D crystals

21. 2D crystals

22. 2D waves

23. Brillouin zone.

24. Non-translational crystal symmetries.

25. Wallpaper groups.

26. Continuous groups.

27. Spherical symmetry.

28. Commutator algebra.

29. Representations of the spherical group.

30. Spherical harmonics.

31. Atomic physics.

32. SU(2) formal and informal, The name of the commutator (Lie) group that describes spherical symmetry is SU(2) the two-dimensional Special Unitary group.

Since we are into matrices of 2-numbers, we need some vocabulary. An *orthogonal* matrix R, such as a rotation, has its transpose as its inverse, $R^T = R^{-1}$, e.g.

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$$

where c is $\cos\theta$ and s is $\sin\theta$.

The 2-number equivalent is the *unitary* matrix U whose *hermitian transpose* $U^\dagger$ is its inverse, $U^\dagger = U^{-1}$ where $(U^\dagger)_{jk} = (U^*)_{kj}$ for the 2-number (complex) conjugate, $*$. E.g.,

$$\begin{pmatrix} c+irs & (ip+q)s \\ (ip-q)s & c-irs \end{pmatrix} \begin{pmatrix} c-irs & -(ip+q)s \\ -(ip-q)s & c+irs \end{pmatrix} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$$

where c is $\cos(\psi/2)$, s is $\sin(\psi/2)$ and the direction of the rotation axis satisfies $p^2 + q^2 + r^2 = 1$.

The 2-number analog of a symmetric matrix ($S^T = S$) is a *hermitian* matrix, $H^\dagger = H$ where, again, ^{T} means transpose, and the hermitian conjugate $^\dagger = ^{T*}$.

Hermitian matrices have symmetric real parts and antisymmetric imaginary parts.

(Note that orthogonal matrices are special cases of unitary matrics, and symmetric matrices are special cases of hermitian matrices.)

Note that the hermitian generators

$$J'_x = i \begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \quad J'_y = i \begin{pmatrix} & 1 \\ -1 & \end{pmatrix} \quad J'_z = i \begin{pmatrix} 1 & -1 \\ & \end{pmatrix}$$

or

$$J_x = \frac{1}{2} \begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 1 & \end{pmatrix} \quad J_y = \frac{1}{2} \begin{pmatrix} & -i & \\ i & & -i \\ & i & \end{pmatrix} \quad J_z = \frac{1}{2} \begin{pmatrix} 1 & & \\ & 0 & \\ & & -1 \end{pmatrix}$$

give rise to unitary group elements $e^{i\vec{\alpha}\cdot\vec{J}}$ since $(e^{i\vec{\alpha}\cdot\vec{J}})^\dagger = e^{-i\vec{\alpha}\cdot\vec{J}^\dagger} = e^{-i\vec{\alpha}\cdot\vec{J}}$ so $(e^{i\vec{\alpha}\cdot\vec{J}})(e^{i\vec{\alpha}\cdot\vec{J}})^\dagger = I$.

That's the "unitary" in SU(2). The "special" means restrict the group elements to matrices whose determinant is 1. In particular, the determinant is not -1, which excludes reflections and inversions from the group.

The group we used in the end to describe atomic symmetry is SO(3), the special *orthogonal* group in three dimensions. This is the subgroup of SU(2) that leaves out the even-dimensional

representations—in which ℓ is half-integral.

For e^{iX} to have determinant $=1$, X must have trace $=0$. (We can see this by imagining coordinate axes in which X is diagonal, so $\det(e^{iX}) = e^{iX_{11}}e^{iX_{22}}..e^{iX_{nn}} = e^{i(X_{11}+X_{22}+..+X_{nn})}$.) Hence traceless generators are always needed for SU groups.

There can only be three independent traceless hermitian generators in two dimensions. Call such a generator

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

where a, b, c and d are 2-numbers. For zero trace $d = -a$. For hermitian, a is real and $c = b^*$. So we are left with

$$\begin{pmatrix} a & b_1 - ib_2 \\ b_1 + ib_2 & -a \end{pmatrix}$$

which has only three numbers to pick. The half-Pauli matrices are one possible choice: $(a, b_1, b_2) = (0, 1/2, 0)$ for $J_x, (0, 0, 1/2)$ for J_y and $(1/2, 0, 0)$ for J_z.

The $1/2$ in the half-Pauli matrices is needed to give the commutators $[J_x, J_y] = iJ_z$, $[J_y, J_z] = iJ_x$ and $[J_z, J_x] = iJ_y$.

The commutators of the raising and lowering operators, once we've switched from J_x and J_y to $J_\pm = (J_x \pm iJ_y)\sqrt{2}$, are more interesting.

$$[J_+, J_-] = J_z \quad [J_z, J_\pm] = \pm J_\pm$$

(Note that J_+ and J_- are no longer hermitian but $J_- = J_+^\dagger$.)

We can see why they are interesting by considering a special representation, the *adjoint representation*. Just as a finite group has a *regular* representation (Note 11) in which the number of dimensions equals the number of group elements, so a Lie group has an adjoint representation, in which the number of dimensions equals the number of generators. (A Lie group representation with one dimension for each element of the group would not be likely to give useful insights.)

For generators $X_1, X_2, .., X_n$ of a Lie group of n generators

$$[X_1, X_2] = i \sum_{k=1}^{n} f_{12k} X_k$$

etc., defines the coefficients f_{ajk}. The n generators of the adjoint representation are the matrices

$$(T_a)_{jk} = -if_{ajk}$$

(note the $-$ sign).

For SU(2) the commutators are

$[J_a, J_j]$		J_x J_1	J_y J_2	J_z J_3
J_x	J_1		iJ_3	$-iJ_2$
J_y	J_2	$-iJ_3$		iJ_1
J_z	J_3	iJ_2	$-iJ_1$	

(I've added the alternative notation $J_x = J_1, J_y = J_2, J_z = J_3$.)

So the f_{ajk}, grouped under $a = 1, 2, 3$, are

$T_1'(a=1)$			
k	1	2	3
j	$\mid J_x >$	$\mid J_y >$	$\mid J_z >$
1 $< J_x \mid$			
2 $< J_y \mid$			$-i$
3 $< J_z \mid$		i	

$T_2'(a=2)$			
k	1	2	3
j	$\mid J_x >$	$\mid J_y >$	$\mid J_z >$
1 $< J_x \mid$			i
2 $< J_y \mid$			
3 $< J_z \mid$	$-i$		

$T_3'(a=3)$			
k	1	2	3
j	$\mid J_x >$	$\mid J_y >$	$\mid J_z >$
1 $< J_x \mid$		$-i$	
2 $< J_y \mid$	i		
3 $< J_z \mid$			

Note first that these are exactly the same as the (spin-1) three-dimensional representation of SU(2) (Note 27), so the "adjoint representation" is indeed a representation.

Note second the defining property of the adjoint representation, that there is one dimension for each group generator, so that we can label the axes $\mid J_x >$, $\mid J_y >$ and $\mid J_z >$ just as well as 1, 2 and 3. This enables us to write notation such as

$$< J_y \mid T_1' \mid J_z > = -i \quad \text{or} \quad T_1' \mid J_z > = -i \mid J_y >$$

for

$$(0,1,0) \begin{pmatrix} & -i & \\ & & \\ & i & \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -i \quad \text{or} \quad \begin{pmatrix} & -i & \\ & & \\ & i & \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -i \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

See how the "bra" $< J_y \mid$ and the "ket" $\mid J_y >$ in the "bracket" $< J_y \mid .. \mid J_y >$ are just hermitian transposes of each other in matrix notation.

We can transform the T' so that one of them becomes diagonal—and we follow our earlier convention in choosing T_z'. We did this in Note 29, but here it is again, keeping track this time of the axes in the bras and kets.

Using $Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i & \\ & & -\sqrt{2} \\ -1 & -i & \end{pmatrix}$ as in Note 29 we change $T' = pT_1' + qT_2' + rT_3'$ to $T = QT'Q^{-1}$.

	$\mid J_+ >$	$\mid J_z >$	$- \mid J_- >$
$< J_+ \mid$	r	$\frac{p-iq}{\sqrt{2}}$	
$< J_z \mid$	$\frac{p+iq}{\sqrt{2}}$		$\frac{p-iq}{\sqrt{2}}$
$- < J_- \mid$		$\frac{p+iq}{\sqrt{2}}$	$-r$

where $\mid J_+ > = \mid J_x + iJ_y > = \mid J_x > +i \mid J_y > = Q^{-1} \mid J_x >$
and $- \mid J_- > = - \mid J_x - iJ_y > = -(\mid J_x > -i \mid J_y >) = Q^{-1} \mid J_x >$.

(Note that the *axes* $\mid J_a >$ transform oppositely to *vectors* of coordinates based on those axes: think of rotations.)

If we further write $T_\pm = (T_1 \pm iT_2)/\sqrt{2}$ (without changing axes again) we have, as in Note 29,

T_+'			
	$\mid J_= >$	$\mid J_z >$	$- \mid J_- >$
$< J_+ \mid$		1	
$< J_z \mid$			1
$- < J_- \mid$			

T_-'			
	$\mid J_+ >$	$\mid J_z >$	$- \mid J_- >$
$< J_+ \mid$			
$< J_z \mid$	1		
$- < J_- \mid$		1	

T_z'			
	$\mid J_+ >$	$\mid J_z >$	$- \mid J_- >$
$< J_+ \mid$	1		
$< J_z \mid$			
$- < J_- \mid$			-1

Now compare the diagonalized adjoint representation with the commutators

$$[J_z, J_z] = 0 \qquad T_z \mid J_z > \ = 0$$

$$[J_z, J_\pm] = \pm J_\pm \qquad T_z \mid J_\pm > \ = \pm J_\pm$$

This is not a coincidence. Remembering that the sum of products

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} (1,0,0,..,0) + \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} (0,1,0,..,0) + .. + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} (0,0,0,..,1) = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

4

we can write (T_3 is the other name for T_z)

$$
\begin{aligned}
T_3 \mid J_a > \; &= \; \sum_b \mid J_b >< J_b \mid T_3 \mid J_a > \\
&= \; \sum_b \mid J_b > (T_3)_{ba} \\
&= \; \sum_b (T_3)_{ba} \mid J_b > \\
&= \; \sum_b -f_{3ba} \mid J_b > \\
&= \; \sum_b f_{3ab} \mid J_b > \\
&= \; \mid \sum_b f_{3ab} J_b > \\
&= \; \mid [J_3, J_a] >
\end{aligned}
$$

so the elements of the diagonal matrix in the adjoint representation are the coefficients of the commutator. Thus if a diagonal matrix D in the adjoint representation were

$$
D = \quad
\begin{array}{c}
\\
< D \mid \\
< X_2 \mid \\
< X_3 \mid \\
\vdots
\end{array}
\begin{array}{cccc}
\mid D > & \mid X_2 > & \mid X_3 > & .. \\
\left| \begin{array}{cccc}
0 & & & \\
& c_2 & & \\
& & c_3 & \\
& & & \ddots
\end{array} \right|
\end{array}
$$

then $[D, X_j] = c_j X_j$.

Note that $\mid D >$ itself is one of the vectors and the diagonal element $< D \mid .. \mid D >$ must be 0 because $[D, D] = 0$.

So the diagonal matrix T_z (or diagonal matrices in general) in the adjoint representation is (are) given directly by the commutators of the corresponding group element with the elements corresponding to the rows of the diagonalizing matrix. For SU(2)—and, we will see, for SU(3)—these "diagonalizing elements" are the raising and lowering operators.

Knowing this, we need not construct the adjoint representation at all, if we are interested only in its diagonal matrices. We can get all their elements directly from the commutators of the group elements.

SU(2) can be thought of variously in two dimensions, in three dimensions, and in one dimension.
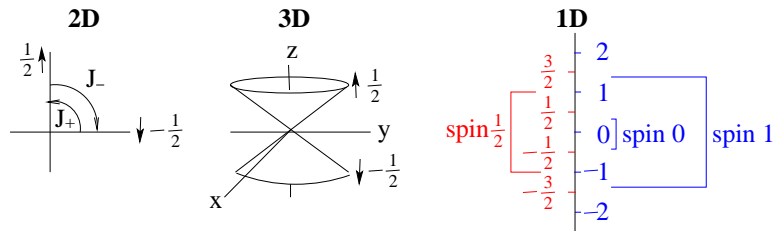
The 2D illustration is the *fundamental representation* of 1/2 spin-up vs. 1/2 spin-down, the smallest representation that distinguishes all the symmetry operations. This is the basic quantum-mechanical space describing the two states of spin 1/2. It has, of course, 2-numbers along each

dimension. States of higher spin, including the 3D 1-number spin 1, are built up from it by tensor products as we explained in Note 29.

The 3D illustration shows the two spin states in the physical world of three dimensions, once we have made the arbitrary choice of diagonalizing in the z direction. Note 31 illustrates this for atoms of integer spin (especially spin 1), but here we show spin 1/2 ($\sqrt{\ell(\ell+1)} = \sqrt{3/2}$).

The 1D illustration shows not just the spin 1/2 representation but the (smaller of) the sets of quantum numbers characterizing all representations. Each set is identified by the total spin: 0, 1/2, 1, 3/2, 2, .. These total spins fall into two categories, the half-integer spins in red and the integer spins in blue. Note that the separation between any two states of any given total spin is always the same value, 1.

33. SU(3). Without at first inquiring into its physical significance, let's explore the next special unitary group, SU(3).

Just as there were three independent SU(2) 2×2 hermitian generator matrics (Note 32), so there are eight independent SU(3) 3×3 matrices.

The hermitian matrix

$$\begin{pmatrix} a & b^* & c^* \\ b & d & e^* \\ c & e & f \end{pmatrix}$$

with a, d and f real, has nine independent numbers, one of which is given by requiring the determinant to be 1 ("special").

The half-Gell-Mann matrices, $X_1, .., X_8$, have useful aliases, $I_x, I_y, I_z, U_x, U_y, V_x, V_y$ and Y, so we show both names. (I_a will come to stand for "isospin" and should not be confused with the identity matrix I.)

$$
\begin{array}{ccccc}
X_1, I_x & X_2, I_y & X_3, I_z & I_+ & I_- \\[2pt]
\frac{1}{2}\begin{pmatrix} & 1 & \\ 1 & & \\ & & \end{pmatrix} &
\frac{1}{2}\begin{pmatrix} & -i & \\ i & & \\ & & \end{pmatrix} &
\frac{1}{2}\begin{pmatrix} 1 & & \\ & -1 & \\ & & \end{pmatrix} &
\frac{1}{\sqrt{2}}\begin{pmatrix} & 1 & \\ & & \\ & & \end{pmatrix} &
\frac{1}{\sqrt{2}}\begin{pmatrix} & & \\ 1 & & \\ & & \end{pmatrix} \\[12pt]
X_4, U_x & X_5, U_y & U_z & U_+ & U_- \\[2pt]
\frac{1}{2}\begin{pmatrix} & & 1 \\ & & \\ 1 & & \end{pmatrix} &
\frac{1}{2}\begin{pmatrix} & & -i \\ & & \\ i & & \end{pmatrix} &
\frac{1}{2}\begin{pmatrix} 1 & & \\ & & \\ & & -1 \end{pmatrix} &
\frac{1}{\sqrt{2}}\begin{pmatrix} & & 1 \\ & & \\ & & \end{pmatrix} &
\frac{1}{\sqrt{2}}\begin{pmatrix} & & \\ & & \\ 1 & & \end{pmatrix} \\[12pt]
X_6, V_x & X_7, V_y & V_z & V_+ & V_- \\[2pt]
\frac{1}{2}\begin{pmatrix} & & \\ & & 1 \\ & 1 & \end{pmatrix} &
\frac{1}{2}\begin{pmatrix} & & \\ & & i \\ & -i & \end{pmatrix} &
\frac{1}{2}\begin{pmatrix} -1 & & \\ & & \\ & & 1 \end{pmatrix} &
\frac{1}{\sqrt{2}}\begin{pmatrix} & & \\ & & \\ & 1 & \end{pmatrix} &
\frac{1}{\sqrt{2}}\begin{pmatrix} & & \\ & & 1 \\ & & \end{pmatrix} \\[12pt]
X_*, Y & & & & \\[2pt]
\frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & & \\ & 1 & \\ & & -2 \end{pmatrix} & & & &
\end{array}
$$

I have shown eight matrices in addition to the eight half-Gell-Mann matrices $X_1, .., X_8$. These will be useful in understanding and working with SU(3) representations.

Note that $I_z + U_z + V_z = 0$ so they are not independent. $Y = (U_z - V_z)/\sqrt{3}$ is independent of I_z. I_z and Y are the two diagonal matrices of the set of eight, and there can be no more than two independent traceless diagonal matrices since $a + b + c = 0$ in

$$\begin{pmatrix} a & & \\ & b & \\ & & c \end{pmatrix}$$

The commutators are

		I_x X_1	I_y X_2	I_z X_3	U_x X_4	U_y X_5	V_x X_6	V_y X_7	Y X_8
I_x	X_1		iX_3	$-iX_2$	$\frac{i}{2}X_7$	$-\frac{i}{2}X_6$	$\frac{i}{2}X_5$	$-\frac{i}{2}X_4$	
I_y	X_2			iX_1	$\frac{i}{2}X_6$	$\frac{i}{2}X_7$	$-\frac{i}{2}X_4$	$-\frac{i}{2}X_5$	
I_z	X_3				$\frac{i}{2}X_5$	$-\frac{i}{2}X_4$	$-\frac{i}{2}X_7$	$\frac{i}{2}X_6$	
U_x	X_4					$\frac{i}{2}(\sqrt{3}X_8+X_3)$	$\frac{i}{2}X_2$	$\frac{i}{2}X_1$	$\frac{i\sqrt{3}}{2}X_5$
U_y	X_5						$-\frac{i}{2}X_1$	$\frac{i}{2}X_2$	$\frac{i\sqrt{3}}{2}X_4$
V_x	X_6							$\frac{i}{2}(\sqrt{3}X_8-X_3)$	$-\frac{i\sqrt{3}}{2}X_7$
V_y	X_7								$\frac{i\sqrt{3}}{2}X_6$
Y	X_8								

Note that in two cases the linear combinations in the results have more than one generator.

From these commutators we could build the 8×8 matrices of the adjoint representation, but we will need only the two generators that have diagonal half-Gell-Mann matrices, I_z and Y. We know from Note 32 that we can get these more directly from the commutators.

	I_+	I_-	I_z	U_+	U_-	V_+	V_-	Y
I_z	I_+	$-I_-$	0	$\frac{1}{2}U_+$	$-\frac{1}{2}U_-$	$-\frac{1}{2}V_+$	$\frac{1}{2}V-$	0
Y	0	0	0	$\frac{\sqrt{3}}{2}U_+$	$-\frac{\sqrt{3}}{2}U_-$	$\frac{\sqrt{3}}{2}V_+$	$-\frac{\sqrt{3}}{2}V_-$	0

We now have two representations for SU(3): the 3-dimensional fundamental representation given by the half-Gell-Mann matrices and the 8-dimensional adjoint representation. Let's see if we can find a third before we go on.

If we replace X_j by $-X_j^*$, the negative of the 2-number (complex) conjugate of every element, we get a new representation because the commutators are the same.

$$[-X_j^*, -X_k^*] = (if_{jk\ell}X_\ell)^* = if_{jkl}(-X_\ell^*)$$

(since the f are always real).

Thus, the 3D fundamental representation gives rise to a different 3D representation with, in particular

$$I_z' = \frac{1}{2}\begin{pmatrix} -1 & & \\ & 1 & \\ & & \end{pmatrix} \quad \text{and} \quad Y' = \frac{1}{2\sqrt{3}}\begin{pmatrix} -1 & & \\ & -1 & \\ & & 2 \end{pmatrix}$$

With these three representations known, we can start trying to classify all representations of SU(3).

Just as SU(2) representations can be classified in one dimension by a single number, the z-component of the total spin, so the SU(3) representations can be classified in two dimensions.

The SU(2) classification is given by the diagonal elements of the one matrix we chose to diagonalize, J_z.

The SU(3) classification is given by the diagonal elements, in pairs, of its two diagonalized matrices, I_z and Y.

Here are the points (it is convenient to write $\sqrt{3}/2$ as $3/(2\sqrt{3})$, etc.).

		3D			3D′			8D						
I_z	$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	1	-1	0	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	0
Y	$\frac{1}{2\sqrt{3}}$	$\frac{1}{2\sqrt{3}}$	$-\frac{2}{2/\sqrt{3}}$	$-\frac{1}{2/\sqrt{3}}$	$-\frac{1}{2\sqrt{3}}$	$\frac{2}{2\sqrt{3}}$	0	0	0	$\frac{3}{2/\sqrt{3}}$	$-\frac{3}{2/\sqrt{3}}$	$\frac{3}{2/\sqrt{3}}$	$-\frac{3}{2/\sqrt{3}}$	0

7
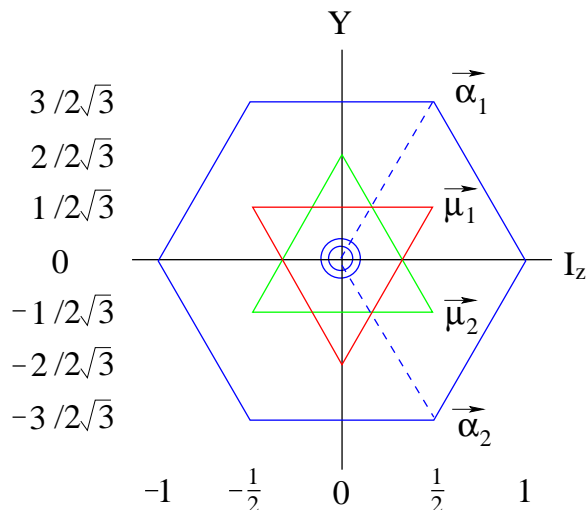
Just as the SU(2) representations occupy two distinct lattices in their one dimension, so the SU(3) representations occupy three distinct lattices in their two dimensions, here distinguished by blue for the adjoint representation, red for the fundamental representation, and green for its conjugate representation.

Just as the spaces between values for any given representation in the one dimension for SU(2) are given by the distance from the origin to the positive value 1 of the adjoint representation, so the spaces between values for any given representation in the two dimensions for SU(3) are given by the distance from the origin to two of the positive values $\vec{\alpha}_1 = (1/2, 3/(2\sqrt{3}))$ and $\vec{\alpha}_2 = (1/2, -3/(2\sqrt{3}))$ of the adjoint representation.

For example, the eight points of the adjoint representation are the origin 0, $\vec{\alpha}_1$, $\vec{\alpha}_2$, $\vec{\alpha}_1 + \vec{\alpha}_2$, $-\vec{\alpha}_1$, $-\vec{\alpha}_2$ and $-\vec{\alpha}_1 - \vec{\alpha}_2$.

The three points of the fundamental representation are $\vec{\mu}_1 = (1/2, 1/(2\sqrt{3}))$, $\vec{\mu}_1 - \vec{\alpha}_1$ and $\vec{\mu}_1 - \vec{\alpha}_1 - \vec{\alpha}_2$. Similarly, its conjugate representation is given by $\vec{\mu}_2 = (1/2, -1/(2\sqrt{3}))$, $\vec{\mu}_2 - \vec{\alpha}_2$ and $\vec{\mu}_2 - \vec{\alpha}_2 - \vec{\alpha}_1$.

Note that all inner points, within the boundaries shown, satisfying these distances from other points in the representation, are also in the representation. In the adjoint representation, this includes the origin, where there are in fact *two* coincident points (for the two diagonal quantities, which of course commute with each other).

The symmetries of the two-dimensional depictions of the representations—the equilateral triangles for the 3D representations and the regular hexagon for the 8D representation—are a justification for the $1/\sqrt{3}$ factor in the Gell-Mann Y matrix definition.

Here is another. We have the relationship

$$I_z + U_z + V_z = 0$$

and

$$Y = \frac{1}{\sqrt{3}}(U_z - V_z)$$

What two-dimensional relationship does this bring to mind? How about
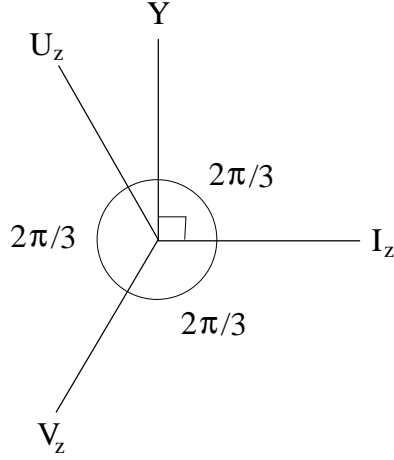
$$1 + e^{2i\pi/3} + e^{4i\pi/3} = 0?$$

Then

$$\frac{1}{\sqrt{3}}(e^{2i\pi/3} - e^{4i\pi/3}) = \frac{1}{\sqrt{3}}(e^{2i\pi/3} - e^{-2i\pi/3}) = \frac{1}{\sqrt{3}}2i\sin\frac{2\pi}{3} = i$$

That is.

SU(3) is made up of three SU(2) spin groups, symmetrically oriented in the plane of the representations.

The three representations we have already have vertices on three distinct grids

$$m(1,0) + n(\frac{1}{2}, \frac{3}{3\sqrt{3}})$$

for the adjoint representation,

$$m(1,0) + n(\frac{1}{2}, \frac{3}{3\sqrt{3}}) - (0, \frac{2}{2\sqrt{3}})$$

for the (red) fundamental representation, and

$$m(1,0) + n(\frac{1}{2}, \frac{3}{3\sqrt{3}}) + (0, \frac{2}{2\sqrt{3}})$$

for the conjugate fundamental representation, all for any integers m and n.

These are the only three possible grids with threefold rotational symmetry about the origin and with edges between nearest-neighbour vertices lying in the I_z, U_z and V_z directions.

Furthermore the dot product between the vertices of any of these grids and the vertices of the grid for the adjoint representation are half-integers.

$$(m + \tfrac{n}{2}, (3n + \left\{ \begin{array}{c} -2 \\ 0 \\ 2 \end{array} \right\})/(2\sqrt{3})) \left(\begin{array}{c} m' + n'/2 \\ (3n' + \left\{ \begin{array}{c} -2 \\ 0 \\ 2 \end{array} \right\})/(2\sqrt{3}) \end{array} \right)$$

$$= mm' + \tfrac{mn' + m'n}{2} + nn' + \left\{ \begin{array}{c} -1 \\ 0 \\ 1 \end{array} \right\} \tfrac{n+n'}{2} + \left\{ \begin{array}{ccc} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{array} \right\} /3$$

$$= mm' + \tfrac{mn' + m'n}{2} + nn' + \left\{ \begin{array}{c} -1 \\ 0 \\ 1 \end{array} \right\} \tfrac{n \text{ or } n'}{2}$$

Where the last follows if we select the zero component, corresponding to the grid of the adjoint representation, for either side.

So the adjoint representation is special. We will call its eight vectors the *roots* of the group SU(3), and we will single out two independent roots spanning the two dimensions. These could be the roots

in the U_z and V_z directions, but we follow others who define the two *simple roots* to be "positive" in the lexicographic sense that their first nonzero components are positive, and to be "simple" in that they are not the sum of any other positive roots.

Positiveness induces a lexicographic ordering on the roots in that roots $\vec{r_1} < \vec{r_2}$ if $\vec{r_2} - \vec{r_1}$ is positive, just as the word `algebra` comes before the word `and` in any dictionary.

Simpleness prevents (1,0) from being a simple root, although it is positive, because

$$(1,0) = (\frac{1}{2}, \frac{3}{2\sqrt{3}}) + (\frac{1}{2}, -\frac{3}{2\sqrt{3}})$$

Finally, the two simple roots for SU(3) are

$$\vec{\alpha_1} = (\frac{1}{2}, \frac{3}{2\sqrt{3}})$$

and

$$\vec{\alpha_2} = (\frac{1}{2}, -\frac{3}{2\sqrt{3}})$$

for any vector $\vec{\mu}$ in any of the three grids and either simple root $\vec{\alpha}$

$$\vec{\alpha} \cdot \vec{\mu}$$

is a half integer.

(More correctly, $\vec{\alpha} \cdot \vec{\mu} / \mid \alpha \mid^2$ is a half integer, but our αs both have length 1.)

There will be some maximum nonnegative integer p such that

$$\vec{\alpha} \cdot (\vec{\mu} + p\vec{\alpha}) = \vec{\alpha} \cdot \vec{\mu} + p = \ell$$

where $2\ell + 1$ is the dimension of the SU(2) representation for U_z and for V_z (and for I_z). There will also be some minimum nonnegative integer q such that

$$\vec{\alpha} \cdot (\vec{\mu} - q\vec{\alpha}) = \vec{\alpha} \cdot \vec{\mu} - q = -\ell$$

This is due to the limits $-\ell, .., \ell$ of the $2\ell + 1$ dimensional spin representation for SU(2) in the directions U_z and V_z (and I_z).

The diagram for the 3D fundamental and the 8D adjoint representations shows this. The maximum extent of the fundamental representation is 2 in each of the three directions, so $\ell = 1/2$ and $2\ell + 1 = 2$. The maximum extent of the adjoint representation is 3 in each direction, so $\ell = 1$ and $2\ell + 1 = 3$.

To find any representation of SU(3) we are looking for sets of grid points with threefold rotational symmetry about the origin which do not extend beyond $\pm\ell$ in each of the three directions.

From the two half-integer rules

$$\begin{aligned} \vec{\alpha} \cdot \vec{\mu} + p &= \ell \\ \vec{\alpha} \cdot \vec{\mu} - q &= -\ell \end{aligned}$$

we have

$$2\vec{\alpha} \cdot \vec{\mu} + p - q = 0$$

i.e.,

$$\vec{\alpha} \cdot \vec{\mu} = (q - p)/2$$

10

and

$$p + q = 2\ell$$

to guide us, where $\vec{\alpha}$ is $\vec{\alpha}_1 = (1/2, 3/(2\sqrt{3}))$ or $\vec{\alpha}_2 = (1/2, -3/(2\sqrt{3}))$.

For the fundamental representation, $\ell = 1/2$, so $p + q = 2\ell = 1$ and (p, q) can only be $(1,0)$ or $(0,1)$. Thus $(q - p)/2$ can only be

p	1	0
q	0	1
$\frac{q-p}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$

Here are $\vec{\alpha}_1 \cdot \vec{\mu}$ on the left and $\vec{\alpha}_2 \cdot \vec{\mu}$ on the right, where the partial row $\vec{\alpha}_i \cdot$ gives the contribution of the first component to the dot product, and the partial column $\vec{\alpha}_i \cdot$ gives the contribution of the second component to the dot product.

The value of the dot product itself is shown beside each grid point.

The red values are those that satisfy $\mid \vec{\alpha}_i \cdot \vec{\mu} \mid \leq \mid \frac{q-p}{2} \mid$ for *both* $\vec{\alpha}_1$ and $\vec{\alpha}_2$. They are just the points we already know to give the fundamental representation.
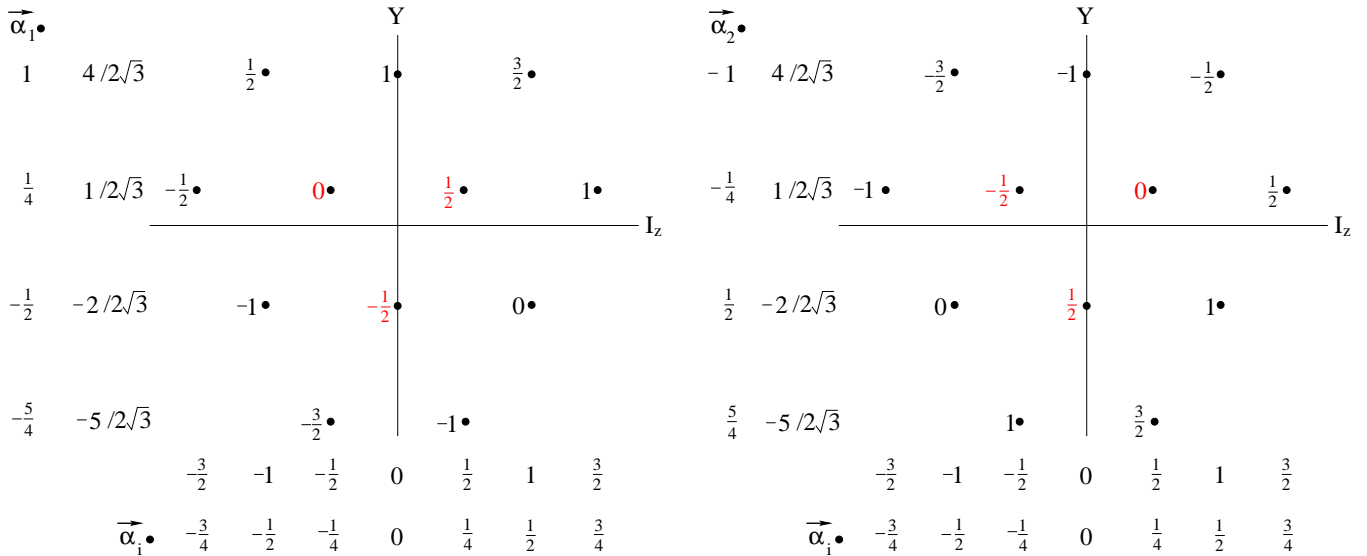

The conjugate fundamental representation is given by the same diagram with the vertical (Y) direction reversed, so we do not need to draw it again.

Even drawing the $\vec{\alpha}_2$ side is not necessary because we can see that it is just the $\vec{\alpha}_1$ side with the horizontal (I_z) direction reversed and the signs changed.

So for the adjoint representation we show only one diagram. Now $\ell = 1$ so $p + q = 2\ell = 1$ allows $(p, q) = (2, 0), (1, 1)$ or $(0,2)$. and

p	2	1	0
q	0	1	2
$\frac{q-p}{2}$	-1	0	1

11
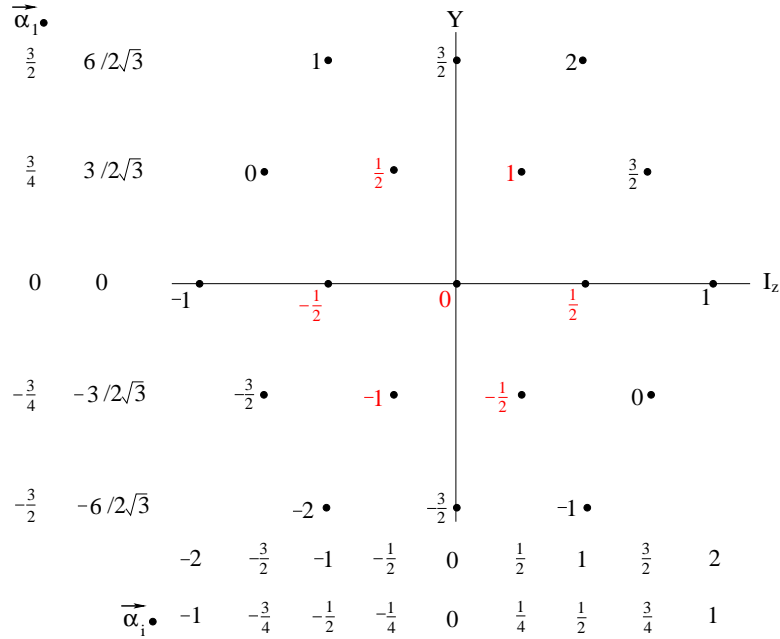
The red values again show points $\vec{\mu}$ where $|\,\vec{\alpha_i} \cdot \vec{\mu}\,|$ for both $\vec{\alpha_1}$ and $\vec{\alpha_2}$, although only $\vec{\alpha_1}$ has been calculated and you must fill in $\vec{\alpha_2}$.

The general form is an "x-u-hexagon" with sides alternating between length x and length u. From the points with the highest Y value and, among these, the point with the highest I_z value, x is the length of the side to it parallel to I_z ("x-axis") and u is the length of the side to it parallel to U_z (the "u-direction").

The fundamental and adjoint representations are the special cases $xu = 10, 01$ and 11, respectively.

Here are some examples, in the grid of the fundamental representation: $xu = 10, 02, 21$ and 13.
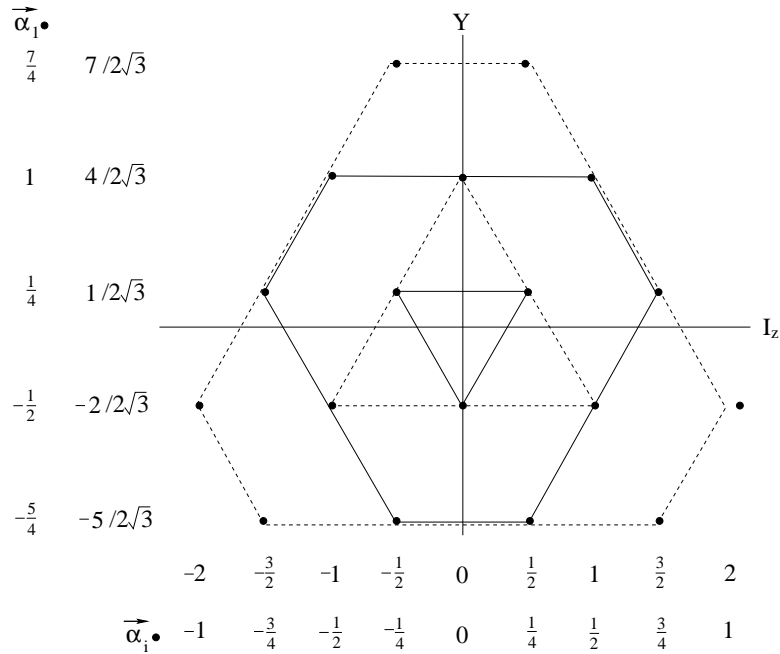
Some of the representations have multiple points: more than one basis vector of the d-dimensional representation maps to the same point in the two-dimensional depiction. For example, the 8D

12

adjoint representation has the six points of the hexagon and a double point at the origin. We will not go into this here, but the total dimension $d = (x + 1)(u + 1)(x + u + 2)/2$.

For triangular representations, $x = 0$ or $u = 0$, there are no multiple points. For regular hexagonal representations, $x = u$, $d = (x + 1)^3$, corresponding to a singleton outer hexagon, a next-inner hexagon of doublets, a next-next-inner hexagon of triplets, and so on.

Here are some numbers

d	$x =$	0	1	2	3	4
$u =$	0	1	3	6	10	15
	1		8	15	24	35
	2			27	42	60
	3				64	90
	4					125

The x and u values directly give the (I_z, Y) coordinates of the rightmost point of the top row of points in the representation. These are

$$\left(\frac{x}{2}, \frac{1}{\sqrt{3}}\left(\frac{x}{2} + u\right)\right)$$

because the highest I_z in the row is $x/2$ by definition of x and $Y = (U_z - V_z)/\sqrt{3} = (I_z + 2U_z)/\sqrt{3}$ using $I_z + U_z + V_z = 0$, and U_z is measured in units of $u/2$.

34. Isospin and quarks. SU(3) is pretty and intriguing in its own right, but it is important for physics as a way of organizing particles smaller than the atom: nucleons and "elementary particles".

The nucleons are the proton p and the neutron n, making up the core or nucleus of the atom. They have nearly identical masses, $m_p = 938.28$MeV and $m_n = 939.57$MeV. There is almost a symmetry here and even in the early 1930s Heisenberg introduced notation for a nucleon doublet

$$\begin{pmatrix} p \\ n \end{pmatrix}$$

which soon was taken up as a depiction of a 2D "spin" state analogous to the spin-up and spin-down electron

$$\begin{pmatrix} \uparrow \\ \downarrow \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \uparrow \\ \rightarrow \end{pmatrix}$$

If there can be doublets corresponding to spin 1/2, there should be triplets corresponding to spin 1. Conveniently, the "pi-mesons" π^+, π^0 and π^- have masses $m_{\pi^0} = 135.0$MeV and $m_{\pi^+} = 139.6$MeV $= m_{\pi^-}$.

So, despite these symmetry-breaking slight differences in mass and, of course, the differences in electron charge (0 for n and π^0, +1 for p and π^+ and -1 for π^-), this property that particles appear in doublets, triplets, and so on, is now known as "isospin". (Originally it was "isobaric spin", referring to baryons, or, misleadingly, "isotopic spin" referring to isotopes.)

SU(2) now has a new application, classifying nucleons and other particles.

But "elementary" particles were being discovered by the dozens, making a disorderly zoo. For instance, there were the $\Lambda^0, \Sigma^+, \Sigma^0, \Sigma^-, \Xi^0$ and Ξ^- particles of masses comparable to the nucleons: 1115.6MeV, 1189.4MeV, 1192.5MeV, 1197.4MeV, 1315MeV and 1321MeV respectively. There also were K^+, K^0, K^- and $\overline{K^0}$ mesons with masses 493.7MeV, 497.7MeV, 493.7MeV and 497.7MeV.

If we include the nucleons in the first group, we have an octet. If we include the π-mesons in the second group we also have almost an octet. Maybe we can find a way to equate these octets with the adjoint representation of SU(3).

To do this requires an interpretation of I_z and Y, characterizing the SU(3) representations. Since

SU(2) symmetry is associated with conservation of angular momentum, J and J_z, maybe I_z and Y are also conserved quantities.

In 1953 Murray Gell-Mann and Kasuohiko Nishijima independently proposed that isospin is conserved in particle reactions involving the "strong" nuclear force.

This would explain the products that arise from collisions of π-mesons and nucleons such as

$$\pi_- + p \longrightarrow \Lambda^0 + K^0$$

or

$$\pi_- + p \longrightarrow \Sigma^- + K^+$$

Looking at the masses (please indulge my great oversimplification of the argument: many other experimental and theoretical considerations must be taken as well) we have isospin singlets, doublets and triplets as follows.

M		I_z
139.6	π^+	1
135.0	π^0	0
139.6	π^-	-1
493.7	K^+	1/2
497.7	K^0	$-1/2$
493.7	$\overline{K^0}$	1/2
497.7	K^-	$-1/2$
938.28	p	1/2
939.57	n	$-1/2$
1115.6	Λ^0	0
1189.4	Σ^+	1
1192.4	Σ^0	0
1197.4	Σ^-	-1

and we see that the combined I_z for $\pi^- + p$ is $-1/2$ which is also the combined I_z for the result. (This is worse than oversimplification. It is downright misleading. Isospins combine in exactly the same way as spins, so we really should work with $I = 3/2$ on both sides and then analyze the components as we did in Note 29. But the gloss speeds up this example.)

If we consider a symmetry-breaking "weak" nuclear force, which does *not* conserve isospin, then we also see why the decays
$$K^0 \longrightarrow \pi^+ + \pi^-$$
and
$$\Lambda^0 \longrightarrow p + \pi^-$$

take so much longer than strong-force processes (10^{-10} sec as opposed to 10^{-23} sec.): they are forbidden to the strong force because isospin is not conserved, but they are permitted by the weak force.

Another quantity which is conserved, not only by strong but by all interactions, is electric charge Q. This is given by the superscripts $+, 0, -$ on the particle names (p is p^+ and n is n^0). For the nucleons
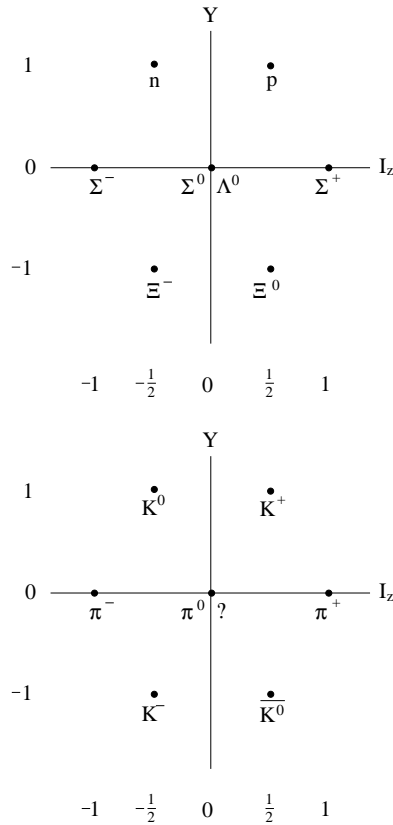$$Q = I_z + \frac{1}{2}$$

but for the other particles we need to invent "hypercharge" Y to generalize this to

$$Q = I_z + \frac{1}{2}Y$$

14

Since Q and I_z are conserved by strong-force interactions, so is Y.

Now we have candidates for both dimensions of the SU(3) representation. Let's see how the particles fit.

The absence of the second (0,0) particle in the meson diagram led to the 1961 discovery of the η^0 particle (549MeV).

One more SU(3) representation appears, involving some particles we have mentioned and which were known in the early 1960s. This is the baryon decuplet
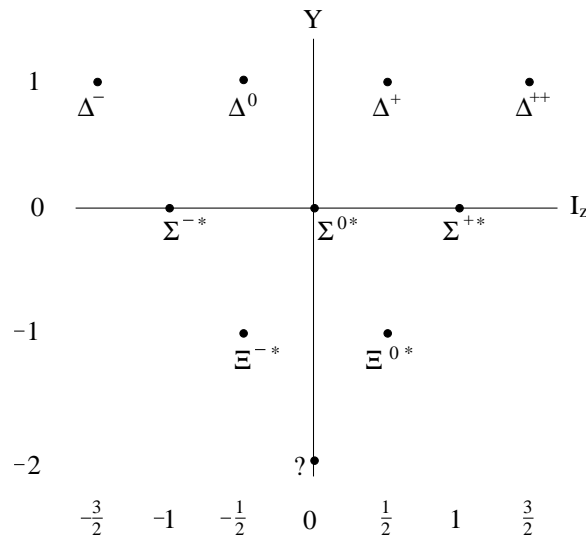
except that one particle was missing, This was found in 1964, the Ω^- particle (1672MeV) with all the right quantum numbers and properties. It was predicted independently by Murray Gell-Mann and Yuval Ne'eman, by using SU(3).

The successes of SU(3) in predicting new particles gave rise to speculations about the fundamental representation, with its three states.

Gell-Mann and George Zwicky independently proposed a triplet of fundamental particles in 1964, which Gell-Mann called "quarks", referring to a line in James Joyce's *Finnigan's Wake*. After a lot of theoretical ramification and limited amounts of experimental evidence, one of the types of quark was observed in 1974, convincing the physics community that the things were more than just theoretical constructs.

The present "standard model" consists of six quarks, from which all "hadrons" (strongly-interacting particles: half-integer spin baryons and integer-spin mesons) are formed, six "leptons" including the electron, and four bosons (integer-spin fundamental particles which mediate the electromagnetic (photon), strong (gluon) and weak (W and Z) forces).

This is better than all the atoms, once considered to be elementary, and better than the particle zoo the standard model emerged from, but still too many (16) particles to be a convincingly final word.

Quarks are more completely described with SU(6) symmetry.

35. **Symmetry and Conservation: Complementary Quantities.** We have seen three pairs of quantities which are complementary in the sense that they appear multiplied together in the exponential 2-number that described the "rotating arrow" (Week 5) that gives frequencies and wavenumbers to particles.

In Week 7a Note 2 we had

$$e^{-i(\omega t - kx)} = e^{-i(Et - px)/\hbar}$$

so that E and t are complementary in this sense, as are p and x.

In Note 26 of this week we had

$$e^{im\phi} = e^{iJ\phi/\hbar}$$

so that J and ϕ are complementary.

Complementary quantities	
time, t	energy, E
space, x	momentum, p
angle, ϕ	angular momentum, J

Furthermore, one quantity of each pair is conserved (in certain circumstances not yet specified) we have claimed: energy, momentum and angular momentum.

This has made working with the conserved quantity of any pair more fruitful than working with its complement: energentum instead of timespace in Week 7a, angular momentum instead of angle in atomic and particle physics.

The complementarity comes out further if we note that

$$\text{slope}_t e^{-i(Et-px)/\hbar} = -\frac{iE}{\hbar}e^{-i(Et-px)/\hbar}$$

$$\text{slope}_x e^{-i(Et-px)/\hbar} = \frac{ip}{\hbar}e^{-i(Et-px)/\hbar}$$

$$\text{slope}_\phi e^{iJ\phi/\hbar} = \frac{iJ}{\hbar}e^{iJ\phi/\hbar}$$

so that we can say

$$
\begin{aligned}
E &= i\hbar \ \text{slope}_t \\
p &= -i\hbar \ \text{slope}_x \\
J &= -i\hbar \ \text{slope}_\phi
\end{aligned}
$$

at least in connection with the 2-number exponential describing the motion of the particle.

Let's be conventional, and save words, by calling this the *wavefunction*:

wavefunction $\Psi(t, x)$ = 2-number function describing the motion of the particle

In the atomic physics of Notes 26 and 31 we found that the probability distribution $e^{im\phi}e^{-im\phi}$ does not depend on the angle ϕ. And we stated that the angular momentum J_z is conserved.

This make us wonder if there is some general function $\mathcal{L}$ whose symmetries in time, space, angular direction, etc. implies conservation of the corresponding complementary quantity.

Since we are dealing with quantities which are perceived to be continuous, at least at the level of the physics we are now considering, so "symmetry" in a quantity such as x means that this function we seek would not explicitly depend on x, and so on.

$$
\begin{aligned}
\text{slope}_x \mathcal{L} = 0 &\Rightarrow \text{Slope}_t p = 0 \\
\text{slope}_\phi \mathcal{L} = 0 &\Rightarrow \text{Slope}_t J = 0 \\
\text{slope}_t \mathcal{L} = 0 &\Rightarrow \text{Slope}_t E = 0
\end{aligned}
$$

where $\mathcal{L}$ is the function to be discovered and the $\Rightarrow \text{Slope}_t q = 0$ expresses that the quantity $q(= p, J, E, ..)$ does not change in time, i.e, is conserved.

The last of these three implications suggests strongly that $\mathcal{L}$ is somehow related to the energy E.

36. Symmetry and Conservation: Energy. So far, the only energy we have seen is E in

$$
E^2 - p^2 c^2 = m^2 c^4
$$

(Week 7a Note 9) which gives approximately

$$
E = mc^2 + mv^2/2
$$

where the second term is the kinetic energy of non-relativistic physics and will be all that we will think about for now.

Kinetic energy is conventionally called $\mathcal{T}$, so

$$
\mathcal{T} = \frac{mv^2}{2}
$$

(We also had, non-relativistically, in Week 7a Note 9, $p = mv$, so we will be able to use

$$
\mathcal{T} = \frac{p^2}{2m}
$$

and

$$
\mathcal{T} = \frac{pv}{2}
$$

during this discussion.)

Kinetic energy is clearly not conserved. It depends on velocity, which can vary.

(Note that velocity is always relative to some specific observer, so $\mathcal{T}$ will be different for observers themselves moving at different velocities. There is no "absolute rest" to measure $\mathcal{T}$ from.)

To speak of conservation of energy we must recognise different forms of energy: kinetic energy (of motion) is one; heat energy (molecular motion inside a body) is another; energy stored in an electromagnetic field is a third; and so on.

The form of energy we need when no heat, electromagnetic fields, etc are involved is more subtle and we could imagine it was invented just to save the idea that energy is conserved.

It is *potential energy* $\mathcal{V}$ (sometimes $\mathcal{U}$). It turns out to be so useful that, energy conservation or not, we need it.

Let's start with a ball of mass m rolling down a hill of height h.
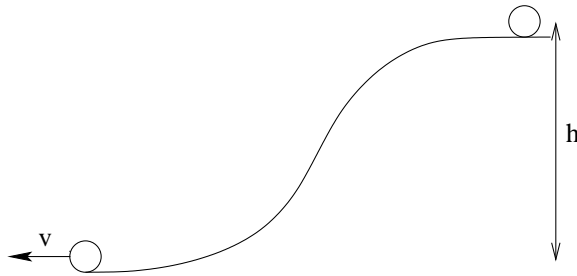
At the bottom it clearly has a velocity v even if it started at the top with zero velocity. Thus, at the bottom $\mathcal{T} = mv^2/2$ while at the top $\mathcal{T} = 0$.

So we say it has potential energy at the top, and we can in fact calculate this

$$\mathcal{V} = mgh$$

where g is a constant specific to gravity at the surface of the Earth

$$g \approx 10 \text{ m/sec}^2$$

Such a potential is called linear: it varies linearly in the space coordinate h. Quadratic and inverse-linear potentials are also common. The excursions include $1/r$ potentials for atoms and solar systems.

Now by saying that energy is conserved we can calculate the velocity of the ball at the bottom in terms of the height at which it started.

$$\frac{mv^2}{2} = mgh$$

so

$$v = \sqrt{2gh}$$

and this relationship between v and h can be confirmed by measuring both.

So $E = \mathcal{T} + \mathcal{V}$ is the total energy (apart from heat, etc.).

Now we are going to show that

$$\mathcal{L} = \mathcal{T} - \mathcal{V}$$

is the symmetry function we postulated in Note 35.

(Note the invertible relationship

$$\begin{pmatrix} E \\ \mathcal{L} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \mathcal{T} \\ \mathcal{V} \end{pmatrix}$$

We can use either $\mathcal{T}$ and $\mathcal{V}$ or E and $\mathcal{L}$.)

How do we get this $\mathcal{L}$? Let's go back to Week 5 Note 7 where we showed that photons travel in straight lines by seeing that their wavefunctions (that is, the 2-number function $c + is = e^{i\omega t}$ that describes the motion of the photon) cancel for all paths except the straight line and those paths extremely close to the straight line.

We were a little careless in Week 5 because we didn't include the distance travelled as well as the time. We should have written, as in Week 7a Note 2

$$e^{-i(\omega t - kx)}$$

(It is a convention to have a negative sign on the $i\omega t$ part, so the arrows rotate clockwise instead of counterclockwise: explaining $c - is$ would have been a distraction then but we need it now.)

Now let's put this in terms of E and p

$$e^{-i(\omega t - kp)} = e^{-i(Et - px)/\hbar}$$

and expand $E = \mathcal{T} + \mathcal{V}$ and $px = pvt = 2\mathcal{T}t$

$$\begin{aligned} e^{-i(\omega t - kp)} &= e^{-i((\mathcal{T}+\mathcal{V})t - 2\mathcal{T}t)/\hbar} \\ &= e^{i\mathcal{L}t/\hbar} \end{aligned}$$

Compare it with $e^{i\omega t}$ from Week 5 and we have $\omega = \mathcal{L}/\hbar$ (*not the same ω as just now*).

What we then did in Week 5 was to insist that the path be *stationary*, i.e., small deviations either side of the path should not change the *total* phase—accumulated from start to end—by much.

We saw that this happens only at the flat part of the added-up total phase angles, and showed the example
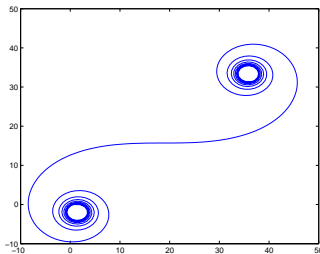
This was easy to work out (with the help of a program) for constant ω, but we know now that $\mathcal{L}$ will vary in space and time so we must get much more sophisticated.

The total phase change is the sum of all the phases as the arrow rotates along its journey from some start time t_1 to the finish time t_2. Since we will think of the rotation and travel happening continuously, that sum becomes the area from t_1 to t_2 under the function $\mathcal{L}$.

We will give some examples in the next Note. In the Week 12 excursion, *Calculus*, we will see that areas are given, remarkably, by antislopes. For the moment, though, we'll be conventional and write it as

$$\int_{t_1}^{t_2} \mathcal{L} dt$$

19

or, showing that $\mathcal{L}$ can depend on t explicitly, as well as on position x and velocity $\dot{c}$,

$$S = \int_{t_1}^{t_2} \mathcal{L}(x, \dot{x}, t)dt$$

This area is called the *action*, S.

We will be looking for a *stationary* action: in principle we want to compare the action for every possible path and select only the path $f(t)$ such that adding a small difference $\eta(t) = f(t) - f'(t)$ to shift to a neighbouring path $f'(t)$ changes the action at most by amounts proportional to η^2 but not proportional to η.

This is called the "principle of least action". It should be called the principle of stationary action: least is indeed stationary, but so is greatest, and even an in-between situation where the action is on a point of inflection
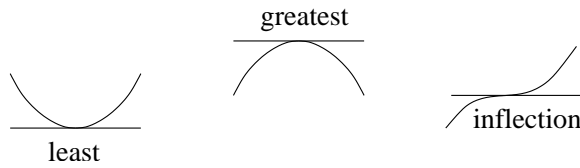
37. **Principle of Stationary Action.** Feynman [FLS64, Ch.19] shows very nicely a special case of the result that stationarity of the action gives

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}} = \frac{\partial \mathcal{L}}{\partial x}$$

or, in our notation

$$\text{Slope}_t \, \text{slope}_{\dot{x}}\mathcal{L} = \text{slope}_x\mathcal{L}$$

the Euler-Lagrange equations, which in turn gives Newton's second law, $F = ma$.

I would like to explore some examples of the "Lagrangian" $\mathcal{L}$ first, and then develop the calculus of variations as an elaboration of the discussion of slopes in Note 30 and in anticipation of the *Calculus* excursion in Week 12.

For the example, let's consider a mass on a spring. We'll hang it up and let it bounce under gravity, although gravity has nothing to do with the problem beyond stretching the spring to its equilibrium point once we first hang it up. We'll call x the vertical displacement from this equilibrium, and $\dot{x}$ the velocity of the mass.

Once the mass and spring are in equilibrium with gravity, we start the process by pulling down on the mass and then letting go.

You know that the mass will rise, go through the equilibrium point, stop, fall back, and repeat. If there were no friction the bouncing would go on forever, so let's pretend there is no friction.

For this oscillating to happen, the potential energy must be in the form of a U (which is why $\mathcal{U}$ is sometimes used to describe it). Let's suppose the U is a parabola. (It actually is, under ideal conditions. When it is, the system is called a harmonic oscillator. When it is not a parabola, it is called an anharmonic oscillator.)

This means

$$\mathcal{V} = \frac{kx^2}{2}$$

and, as always,

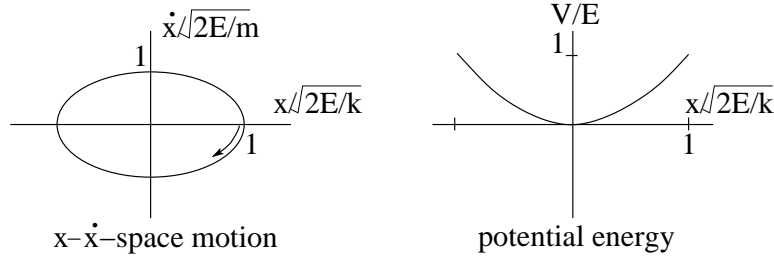$$\mathcal{T} = \frac{m\dot{x}^2}{2}$$

where we now use a frequent convention $\dot{x} = \text{Slope}_t x$.

The total energy

$$E = \mathcal{T} + \mathcal{V} = \frac{m\dot{x}^2}{2} + \frac{kx^2}{2}$$

and, if we acknowledge that E is conserved (we will eventually be able to prove that from the principle of stationary action) we see that the relationship between $\dot{x}$ and x must be an ellipse

$$\frac{\dot{x}^2}{\frac{2E}{m}} + \frac{x^2}{\frac{2E}{k}} = 1$$

x–ẋ–space motion   potential energy

Since we will need the Lagrangian as a function of time, we must find out how x and $\dot{x}$ behave in time.

We can always parameterize the ellipse

$$\frac{\dot{x}^2}{\frac{2E}{m}} + \frac{x^2}{\frac{2E}{k}} = 1$$

as

$$\dot{x} = -\sqrt{\frac{2E}{m}}\sin\theta$$

$$x = \sqrt{\frac{2E}{k}}\cos\theta$$

so that θ is 0 when the motion starts at $(x, \dot{x}) = (\sqrt{2E/k}, 0)$ and where $\dot{x}$ starts to increase in the negative direction of x.

Let's say $\theta = \sqrt{k/m}\, t$. This gives the velocity

$$\dot{x} = \text{Slope}_t x = -\sqrt{\frac{2E}{k}}\sqrt{\frac{k}{m}}\sin\sqrt{\frac{k}{m}}t$$

which is right.

The Lagrangian for the ellipse

$$\mathcal{L} = \mathcal{T} - \mathcal{V} = \frac{m\dot{x}^2}{2} - \frac{kx^2}{2}$$

so

$$\frac{\mathcal{L}}{E} = \sin^2\sqrt{\frac{k}{m}}t - \cos^2\sqrt{\frac{k}{m}}t = -\cos 2\sqrt{\frac{k}{m}}t$$

21

Let's define two useful periods T and t_2. For a complete cycle, $T = 2\pi\sqrt{m/k}$, and for a quarter cycle, $t_2 = (\pi/2)\sqrt{m/k}$.

Let's calculate the Lagrangian for a second path, this time one that does not conserve energy.

$$x(t) = \sqrt{\frac{2E}{k}}(1 - f(t))$$

$$\dot{x}(t) = -\sqrt{\frac{2E}{k}}\dot{f}(t)$$

must satisfy $(t_2 = (\pi/2)\sqrt{m/k})$

$$x(0) = \sqrt{\frac{2E}{k}} \qquad \dot{x}(0) = 0$$

$$x(t_2) = 0 \qquad \dot{x}(t_2) = \sqrt{\frac{2E}{m}}$$

i.e.,

$$f(0) = 0 = \dot{f}(0)$$

$$f(t_2) = 1 \quad \dot{f}(t_2) = \sqrt{\frac{k}{m}}$$

These four conditions can be met if $f(t)$ is a cubic polynomial, say

$$f(t) = z_1 + z_2\frac{t}{t_2} + a(\frac{t}{t_2})^2 + b(\frac{t}{t_2})^3$$

(and $z_1 = 0 = z_2$ because $f(0) = 0 = \dot{f}(0)$).

Such a polynomial, fitted to the two conditions on f and the two conditions on its slope $\dot{f}$ is called a cubic spline. It works out to give almost, but not quite, the same results as the ellipse.

So let's make another condition, say that the x-$\dot{x}$ path passes through the halfway point at $t = t_2/2$.

$$x(\frac{t_2}{2}) = \sqrt{\frac{E}{2k}} \quad \dot{x}(\frac{t_2}{2}) = -\sqrt{\frac{E}{2m}}$$

i.e.,

$$f(\frac{t_2}{2}) = \frac{1}{2} \quad \dot{f}(\frac{t_2}{2}) = \frac{1}{2}\sqrt{\frac{k}{m}}$$

We now have six conditions, and so must fit a quintic spline

$$f(t) = a(\frac{t}{t_2})^2 + b(\frac{t}{t_2})^3 + c(\frac{t}{t_2})^4 + d(\frac{t}{t_2})^5$$

$$t_2\dot{f}(t) = 2a(\frac{t}{t_2}) + 3b(\frac{t}{t_2})^2 + 4c(\frac{t}{t_2})^3 + 5d(\frac{t}{t_2})^4$$

(Use the chain rule for slopes—Note 30—to get $\dot{f}$.)

The four conditions (apart from $f(0) = 0 = \dot{f}(0)$) give

$$\begin{pmatrix} 1 \\ t_2\sqrt{\frac{k}{m}} \\ \frac{1}{2} \\ \frac{t_2}{2}\sqrt{\frac{k}{m}} \end{pmatrix} = \begin{pmatrix} 1 \\ \pi/2 \\ 1/2 \\ \pi/4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \\ 1/4 & 1/8 & 1/16 & 1/32 \\ 1 & 3/4 & 1/2 & 5/16 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

22

which can be solved for the four coefficients.

Here are plots from a MATLAB program to calculate x and $\dot{x}$ for both the ellipse and the new path through the halfway point. I also calculated the Lagrangian $\mathcal{L}$ and the "Hamiltonian" $\mathcal{H}$ (which we can for now take to be just the energy E) for both paths. The program plots, for the first quarter of the ellipse only, since the remaining quarters are closely similar,

- $\dot{x}$ vs x to show the path in x-$\dot{x}$ space

- $x, \dot{x}$ vs t

- $\mathcal{L}, \mathcal{H}$ vs t
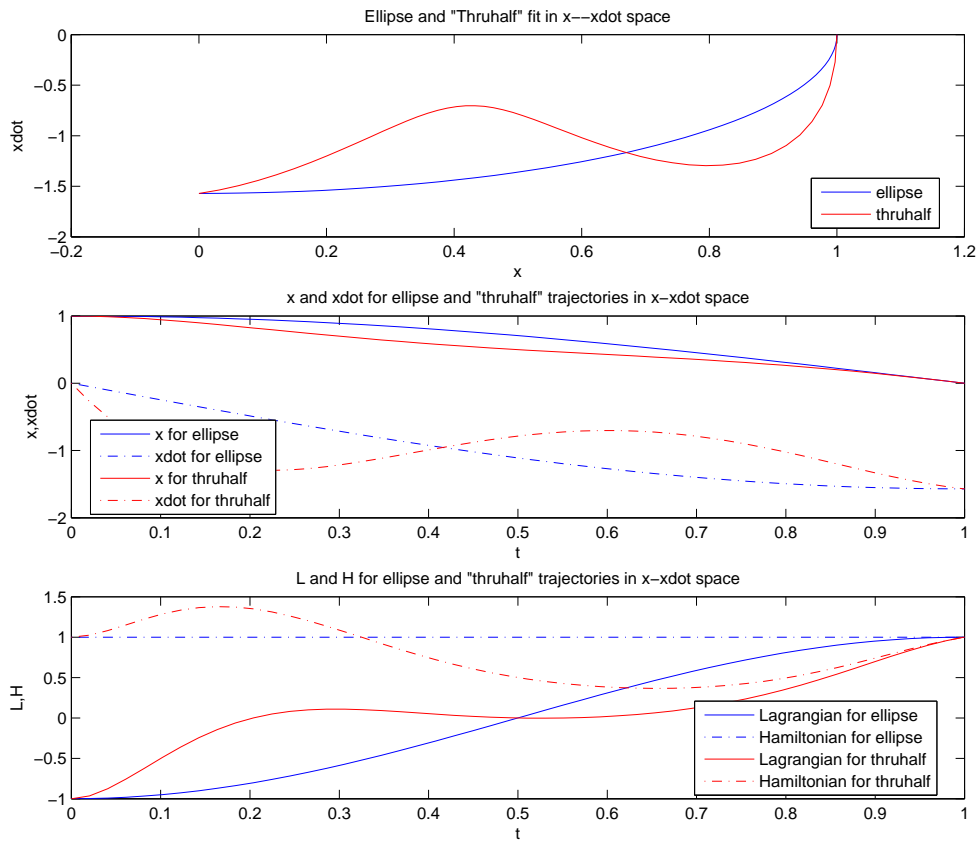
We see that $\mathcal{H}$ 'is a flat line for the ellipse: energy is conserved. But it is not conserved by the path through the halfway point.

The program also calculates the action, a single number for the new path.

The action for the ellipse is 0 as we now see In Note 36 we defined the action

$$S = \int_{t_1}^{t_2} \mathcal{L}(x, \dot{x}, t) dt$$

where $\int_{t_1}^{t_2} \mathcal{L} dt$ is the area under the curve $\mathcal{L}$. Looking at the plot of $\mathcal{L}$ for the ellipse, we see that it

is symmetrical above and below $\mathcal{L} = 0$, so $S = 0$.

Note 36 claims that areas are antislopes (see excursion *Calculus* in Week 12) so let's find antislopes.

For the ellipse

$$\frac{\mathcal{L}(t)}{E} = -\cos 2\sqrt{\frac{k}{m}}t$$

so, apart from an additive constant (see *Calculus*),

$$\frac{\text{antislope}_t \mathcal{L}(t)}{E} = -\sin 2\sqrt{\frac{k}{m}}t$$

and

$$\frac{S}{E} = -\sin 2\sqrt{\frac{k}{m}}t_2 - (-\sin 2\sqrt{\frac{k}{m}}0) = 0$$

by evaluating $\text{antislope}_t \frac{\mathcal{L}}{E}$ at the two endpoints $t = 0$ and $t = t_2 = (\pi/2)\sqrt{k/m}$. (We can write this

$$\text{antislope}_t \frac{\mathcal{L}}{E} \big|_0^{t_2} \ .)$$

For the new path through the halfway point we must work out a lot of coefficients using $\text{antislope}_t t^n = (t^{n-1})/n$.

The kinetic energy

$$\mathcal{T} = \frac{m\dot{x}^2}{2} = \frac{Em}{kt_2^2}(2a(\frac{t}{t_2}) + 3b(\frac{t}{t_2})^2 + 4c(\frac{t}{t_2})^3 + 5d(\frac{t}{t_2})^4)^2$$

where

$$\frac{Em}{kt_2^2} = \frac{4E}{\pi^2}$$

and the potential energy

$$\mathcal{V} = \frac{kx^2}{2} = E(1 - a(\frac{t}{t_2})^2 - b(\frac{t}{t_2})^3 - c(\frac{t}{t_2})^4 - d(\frac{t}{t_2})^5)^2$$

So $S/E = \text{antislope}_t(\mathcal{T} - \mathcal{V}) \big|_0^{t_2}$ is a long expression.

We can work out this expression systematically by recording the coefficients for each power of t/t_2. Let's write $c_n(t/t_2)^n$ as the pair (n, c_n) and work these out first for $\mathcal{T}$ then for $\mathcal{V}$.

$\frac{\pi^2}{4}\frac{\mathcal{T}}{E}$		1	2	3	4
		$2a$	$3b$	$4c$	$5d$
1	$2a$	$2, 4a^2$	$3, 12ab$	$4, 16ac$	$5, 20ad$
2	$3b$		$4, 9b^2$	$5, 24bc$	$6, 30bd$
3	$4c$			$6, 16c^2$	$7, 40cd$
4	$5d$				$8, 25d^2$

$\frac{\mathcal{V}}{E}$		0	2	3	4	5
		1	$-a$	$-b$	$-c$	$-d$
0	1	$0,1$	$2, -a$	$3, -b$	$4, -c$	$5, -d$
2	$-a$		$4, a^2$	$5, ab$	$6, ac$	$7, ad$
3	$-b$			$6, b^2$	$7, bc$	$8, bd$
4	$-c$				$8, c^2$	$9, cd$
5	$-d$					$10, d^2$

(Note that I have written the off-diagonal entries only once, but doubled their coefficients.)

From these tables we can gather coefficients of like powers. Thus, e.g.,

$$c_4 = \frac{4}{\pi^2}(16ac + 9b^2) - (-c + a^2)$$

and so on, as appear in the program (see MATLABpak: SHOquarter.m).

Thus the antislopes, e.g.,

$$\text{antislope}_t c_4 (\frac{t}{t_2})^4 = t_2 \frac{c_4}{5}(\frac{t}{t_2})^5$$

Finally, evaluating at $t = t_2$ and $t = 0$ and subtracting, e.g.,

$$t_2 \frac{c_4}{5}(\frac{t}{t_2})^5 \big|_0^{t_2} = t_2 \frac{c_4}{5}$$

The result is $S/(t_2 E) = -41.48$.

This number is smaller than $S/(t_2 E) = 0$ for the ellipse, which is the true path: the true path certainly does not have least action, which is why the principle is *stationary* action.

These two examples of the Lagrangian and its area, which is the action, help us familiarize ourselves with but do not establish the principle of stationary action.

For that, we need some elaborations on the notion of slopes.

As an example to work with, let's pretend the Lagrangian for the simple harmonic oscillator were an explicit function of time, say

$$\mathcal{L}(t, x, \dot{x}) = \frac{m\dot{x}^2}{2} - \frac{kx^2}{2} + 2Et$$

(This could be obtained if both kinetic and potential energy had a t term but with opposite signs

$$\begin{aligned}\mathcal{T} &= \frac{m\dot{x}^2}{2} + Et \\ \mathcal{V} &= \frac{kx^2}{2} - Et\end{aligned}$$

The total energy $\mathcal{T} + \mathcal{V}$ is still independent of t and so conserved.)

Now we distinguish between a *partial slope*, which we have been using so far

$$\text{slope}_t \mathcal{L} = 2E$$

and a *total slope* which is new and which we write with a capital S

$$\text{Slope}_t \mathcal{L} = 2E + \text{slope}_{\dot{x}} \mathcal{L} \, \text{slope}_t \dot{x} + \text{slope}_x \mathcal{L} \, \text{slope}_t x$$

Here we use the chain rule (Note 30) to extract the total dependence of $\mathcal{L}$ on t, not only explicitly as given by $\text{slope}_t \mathcal{L}$ but also via the t-dependence of any other quantities $\mathcal{L}$ is written in terms of.

Then for the ellipse where

$$\begin{aligned}x &= \sqrt{\frac{2E}{k}} \cos \sqrt{\frac{k}{m}} t \\ \dot{x} &= -\sqrt{\frac{2E}{m}} \sin \sqrt{\frac{k}{m}} t\end{aligned}$$

25

$$\ddot{x} \;=\; \text{Slope}_t\dot{x} = -\frac{\sqrt{2Ek}}{m}\cos\sqrt{\frac{k}{m}}t = -\sqrt{\frac{k}{m}}x$$

$$\text{Slope}_t\mathcal{L} \;=\; 2E + m\dot{x}\ddot{x} - kx\dot{x}$$

$$=\; 2E(1 + 2\sqrt{\frac{k}{m}}\cos\sqrt{\frac{k}{m}}t\sin\sqrt{\frac{k}{m}}t)$$

And now forget about our fable that $\mathcal{L}$ depends explicitly on t. There is still a difference

$$\text{slope}_t\mathcal{L}(x,\dot{x}) \;=\; 0$$

$$\text{Slope}_t\mathcal{L}(x,\dot{x}) \;=\; 4E\sqrt{\frac{k}{m}}\cos\sqrt{\frac{k}{m}}t\sin\sqrt{\frac{k}{m}}t$$

For a function of a single quantity, there is no difference between the total and partial slopes, e.g.,

$$\text{slope}_t x = \dot{x} = \text{Slope}_t x$$

Note that whenever there *is* a difference, the dot notation means total slope with respect to t time

$$\dot{x} \stackrel{\text{def}}{=} \text{Slope}_t x$$

Both total and partial slopes are defined in terms of small changes in the independent variable.

$$\text{Slope}_t\mathcal{L} \;\approx\; \frac{1}{\Delta t}(\mathcal{L}(t+\Delta t, x(t+\Delta t), \dot{x}(t+\Delta t)) - \mathcal{L}(t, x(t), \dot{x}(t)))$$

$$\text{slope}_t\mathcal{L} \;\approx\; \frac{1}{\Delta t}(\mathcal{L}(t+\Delta t, x(t), \dot{x}(t)) - \mathcal{L}(t, x(t), \dot{x}(t)))$$

These definitions give us directly

$$\mathcal{L}(t+\Delta t) \approx \mathcal{L}(t) + \Delta t\,\text{slope}_t\mathcal{L}$$

A useful trick in dealing with antislopes is "partial integration", which follows from the product rule (Note 26) for both partial and total slopes.

$$\text{Slope}_t(f\eta) \;=\; (\text{Slope}_t f)\eta + f(\text{Slope}_t\eta)$$

$$=\; \dot{f}\eta + f\dot{\eta}$$

So

$$\text{antislope}_t(f\dot{\eta}) \;=\; \text{antislope}_t(\text{Slope}_t(f\eta) - \dot{f}\eta))$$

$$=\; f\eta - \text{antislope}_t(\dot{f}\eta)$$

Sometimes the second antislope is easier to work with than the first.

Here is the antislope rule we've just shown, and for convenience here too is the antislope rule that relates to the chain rule (Note 30). We'll abbreviate $g'(x) = \text{slope}_x g(x)$. So, apart from an additive constant, $g(x) = \text{antislope}_x g'(x)$.

slope rule	antislope rule
product	$\text{antislope}_x(f\;\text{slope}_x g) = fg - \text{antislope}_x(g\;\text{slope}_x f)$
chain	$\text{antislope}_f g'(f) = \text{antislope}_x(g'(f)\;\text{slope}_x f(x))$

Now let's look at the action along two different paths. In the above plot of x vs. t (either for ellipse or thruhalf), call the blue path

$$x(t)$$

and call the red path

$$X(t)$$

Call their difference another path $\eta(t)$ so

$$X(t) - x(t) = \eta(t)$$

The two actions are

$$S_{X(t)} = \text{antislope}_t(\mathcal{T}(\dot{X}) - \mathcal{V}(X)) \, |_{t_1}^{t_2}$$

and

$$S_{x(t)} = \text{antislope}_t(\mathcal{T}(\dot{x}) - \mathcal{V}(x)) \, |_{t_1}^{t_2}$$

So the difference

$$
\begin{aligned}
S_{\eta(t)} &= \text{antislope}_t(\mathcal{T}(\dot{x} + \dot{\eta}) - \mathcal{T}(\dot{x}) - \mathcal{V}(x + \eta) + \mathcal{V}(x)) \, |_{t_1}^{t_2} \\
&= \text{antislope}_t(\dot{\eta} \, \text{slope}_{\dot{x}}\mathcal{T} - \eta \, \text{slope}_x\mathcal{V}) \, |_{t_1}^{t_2} \\
&= \text{antislope}_t(\text{Slope}_t(\eta \, \text{slope}_{\dot{x}}\mathcal{T}) - \eta \, \text{Slope}_t(\text{slope}_{\dot{x}}\mathcal{T}) - \eta \, \text{slope}_x\mathcal{V}) \, |_{t_1}^{t_2} \\
&= (\eta \, \text{slope}_{\dot{x}}) \, |_{t_1}^{t_2} - \text{antislope}_t(\eta(\text{Slope}_t(\text{slope}_{\dot{x}}\mathcal{T}) + \text{slope}_x\mathcal{V})) \, |_{t_1}^{t_2}
\end{aligned}
$$

Now we can say three things. First, $\eta(t_1) = 0 = \eta(t_2)$ or else the paths $x(t)$ and $X(t)$ would not start and end at the same time, which is the whole point.

Second, the condition for stationarity is

$$0 = S_{\eta(t)} = \text{antislope}_t(\eta(\text{Slope}_t(\text{slope}_{\dot{x}}\mathcal{T}) + \text{slope}_x\mathcal{V})) \, |_{t_1}^{t_2}$$

for small differences in the path $\eta(t) = X(t) - x(t)$. Since $\eta(t)$ can be *anything* small (as long as $\eta(t_1) = 0 = \eta(t_2)$), this requires

$$\text{Slope}_t(\text{slope}_{\dot{x}}\mathcal{T}) + \text{slope}_x\mathcal{V} = 0$$

Third, we have assumed that $\mathcal{T}$ is not a function of x and $\mathcal{V}$ is not a function of $\dot{x}$. (These assumptions do not always hold.) So we can replace $\mathcal{T}$ by $\mathcal{L}$ and $\mathcal{V}$ by $-\mathcal{L}$ in these slopes:

$$\text{Slope}_t(\text{slope}_{\dot{x}}\mathcal{L}) = \text{slope}_x\mathcal{L}$$

This is the Euler-Lagrange equation we set out to prove at the beginning of this Note.

We can interpret $\text{slope}_{\dot{x}}\mathcal{L}$ as p, the momentum, and $\text{slope}_x\mathcal{L}$ as F, the force on the object, so the Euler-Lagrange equation

$$\text{Slope}_t\text{slope}_{\dot{x}}\mathcal{L} = \text{slope}_x\mathcal{L}$$

says the rate of change of momentum equals the force.

For the simple harmonic oscillator, a mass on a spring,

$$
\begin{aligned}
\mathcal{L} &= \frac{m\dot{x}^2}{2} - \frac{kx^2}{m} \\
\text{slope}_{\dot{x}}\mathcal{L} &= m\dot{x} = p
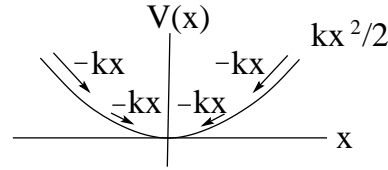\end{aligned}
$$

so

$$\text{Slope}_t\text{slope}_{\dot{x}}\mathcal{L} = m\ddot{x} = \dot{p}$$

Also

$$\text{slope}_x\mathcal{L} = -kx$$

and the force is downhill in the parabolic "potential well"

Potential energy and force

We can take $F = -\text{slope}_x \mathcal{V}(x)$ to be the definition of force that captures our intuitive notion of what we feel when we start rolling downhill or when we squeeze or stretch a spring.

Conservation of energy follows from the Euler-Lagrange equations, as we shall see in Note 38.

Thus the ellipse in x-$\dot{x}$ space is the correct path for the simple harmonic oscillator, while the path going through $(x, \dot{x}) = (\sqrt{E/(2k)}, \sqrt{E/(2m)})$ is not.

38. Symmetry and Conservation: Noether's Theorem. The discussion in Notes 36 and 37 of energy and the principle of least action were a necessary part of our route to showing that the Lagrangian $\mathcal{L}$ is the (actually, a) function whose symmetries give conservation laws.

Conservation of energy $E = \mathcal{T} + \mathcal{V} = 2\mathcal{T} - \mathcal{L} = \dot{x}p - \mathcal{L}$

$$
\begin{aligned}
\dot{E} &= \text{Slope}_t E = \ddot{x}p + \dot{x}\dot{p} - \text{Slope}_t \mathcal{L} \\
&= \ddot{x}p + \dot{x}\dot{p} - \text{slope}_x \mathcal{L}\, \text{Slope}_t x - \text{slope}_{\dot{x}}\mathcal{L}\, \text{Slope}_t \dot{x} \\
&= \ddot{x}p + \dot{x}\dot{p} - \dot{x}\text{slope}_x \mathcal{L} - \ddot{x}\text{slope}_{\dot{x}}\mathcal{L}
\end{aligned}
$$

Since $\text{slope}_{\dot{x}}\mathcal{L} = p$ and since Euler-Lagrange says $\text{slope}_x \mathcal{L} = \text{Slope}_t \text{slope}_{\dot{x}} = \dot{p}$, we can go on

$$
\dot{E} = \ddot{x}p + \dot{x}\dot{p} - \dot{x}\dot{p} - \ddot{x}p = 0
$$

So energy does not change in time: it is conserved.

Conservation of momentum follows directly from any Lagrangian with no explicit x dependence: if

$$
0 = \text{slope}_x \mathcal{L} = \text{Slope}_t \text{slope}_{\dot{x}}\mathcal{L} = \text{Slope}_t p
$$

Thus $\dot{p} = 0$ so p is conserved.

Let's try angular momentum and its complementary quantity, the angle.

In Note 30 we found

$$
\text{Slope}_\phi = x\, \text{slope}_y - y\, \text{slope}_x
$$

and applying these operators to $e^{iJ\phi/\hbar}$ and $e^{i(p_x y + p_y x)/\hbar}$ respectively we can say that angular momentum

$$
J = xp_y - yp_x
$$

in terms of linear momentum in two dimensions $p = p_x e_1 + p_y e_2$.

Back to the Lagrangian: the reasoning is similar to that for x and p above. Suppose

$$
\begin{aligned}
0 &= \text{slope}_\phi \mathcal{L} \\
&= \text{slope}_x \mathcal{L}\, \text{slope}_\phi x + \text{slope}_y \mathcal{L}\, \text{slope}_\phi y + \text{slope}_{\dot{x}}\mathcal{L}\, \text{slope}_\phi \dot{x} + \text{slope}_{\dot{y}}\mathcal{L}\, \text{slope}_\phi \dot{y}
\end{aligned}
$$

We can show from $x = r\cos\phi$ and $y = r\sin\phi$ that $\text{slope}_\phi x = -y$, $\text{slope}_\phi y = x$, $\text{slope}_\phi \dot{x} = -\dot{y}$ and $\text{slope}_\phi \dot{y} = \dot{x}$.

The Euler-Lagrange equation becoes two in two dimensions:

$$
\begin{aligned}
\dot{p}_x &= \text{Slope}_t \text{slope}_{\dot{x}}\mathcal{L} = \text{slope}_x \mathcal{L} \\
\dot{p}_y &= \text{Slope}_t \text{slope}_{\dot{y}}\mathcal{L} = \text{slope}_y \mathcal{L}
\end{aligned}
$$

Putting these together we can confirm

$$
\begin{aligned}
0 = \text{slope}_\phi \mathcal{L} &= -y\,\text{slope}_x\mathcal{L} + x\,\text{slope}_y\mathcal{L} - \dot{y}\,\text{slope}_{\dot{x}}\mathcal{L} + \dot{x}\,\text{slope}_{\dot{y}}\mathcal{L} \\
&= -y\dot{p}_x + x\dot{p}_y - \dot{y}p_x + \dot{x}p_y \\
&= \text{Slope}_t(xp_y - yp_x) \\
&= \text{Slope}_t J
\end{aligned}
$$

Thus $\dot{J} = 0$ so J is conserved.

The neat thing about the Lagrangian formulation is that we can work directly with angles and angular momentum.

$$
\begin{aligned}
J &= \text{slope}_{\dot{\phi}}\mathcal{L} \\
\dot{J} &= \text{Slope}_t\text{slope}_{\dot{\phi}}\mathcal{L} = \text{slope}_\phi\mathcal{L}
\end{aligned}
$$

Let's check this with a two-dimensional simple harmonic oscilltor

$$
\begin{aligned}
\mathcal{L} &= \frac{m(\dot{x}^2 + \dot{y}^2)}{2} - \frac{k(x^2 + y^2)}{2} \\
&= \frac{m(\dot{r}^2 + r^2\dot{\phi}^2)}{2} - \frac{kr^2}{2} \\
J &= \text{slope}_{\dot{\phi}}\mathcal{L} = mr^2\dot{\phi} \\
\dot{J} &= \text{slope}_\phi\mathcal{L} = 0
\end{aligned}
$$

(Compare this with the excursion *Kepler I.*)

Since the Lagrangian is just a number, not a vector, it is independent of the coordinate system, so the Euler-Lagrange equations hold in any coordinate system: (x, y), (x', y') or even polar coordinates (r, ϕ).

Thus the Lagrangian formalism is usually written in terms of "generalized coordinates" q_j and "generalized momenta" p_j. These could be Cartesian coordinates and linear momenta

$$(x, p_x) \quad \text{and} \quad (y, p_y)$$

or polar coordinates and radial and angular momenta

$$(r, p_r) \quad \text{and} \quad (\phi, J)$$

and so on.

In general

$$\text{Slope}_t\,\text{slope}_{\dot{q}_j}\mathcal{L} = \text{slope}_{q_j}\mathcal{L}$$

for each dimension $j = 1, .., d$.

39. The Hamiltonian and Schrödinger's Equation. The Lagrangian formulation uses velocities explicitly. In quantum mechanics, velocity is not properly defined since knowing the precise position of a particle precludes knowing its momentum (see Week 9) and so precludes knowing its "next" position.

But momenta are well defined, so it is important to rework the Lagrangian formulation into a new formulation which replaces velocities by momenta.

This is named after William Rowan Hamilton (1805–65). The Hamiltonian

$$\mathcal{H} = \mathcal{T} + \mathcal{V} = p\dot{q} - \mathcal{L}$$

satisfies two equations

$$\text{slope}_q \mathcal{H} = -\dot{p}$$
$$\text{slope}_p \mathcal{H} = \dot{q}$$

The first follows from $\text{slope}_q \mathcal{L} = \dot{p}$ since $\mathcal{H}$ does not otherwise depend on q.

The second is given by

$$\text{slope}_p \mathcal{H} = \dot{q} + p\,\text{slope}_p \dot{q} - \text{slope}_{\dot{q}} \mathcal{L}\,\text{slope}_p \dot{q} - \text{slope}_q \mathcal{L}\,\text{slope}_p q - \text{slope}_t \mathcal{L}\,\text{slope}_p t$$
$$= \dot{q}$$

because the $\text{slope}_p q$ terms cancel out and because p and q are independent variables so $\text{slope}_p q = 0$ and t of course does not depend on p so $\text{slope}_p t = 0$.

For the simple harmonic oscillator

$$\mathcal{H} = \frac{p^2}{2m} + \frac{kx^2}{2}$$

so

$$\text{slope}_p \mathcal{H} = \frac{p}{m} = v = \dot{x}$$

and

$$\text{slope}_x \mathcal{H} = kx = -F = -\dot{p}$$

With the Hamiltonian formulation we can do quantum physics. For a free particle, the 2-number describing the motion—its wavefunction— is

$$\Psi(t, x) = e^{-i(Et - px)/\hbar}$$

and, applied to this, energy and momentum become the operators

$$E = i\hbar\,\text{slope}_t$$
$$p = -i\hbar\,\text{slope}_x$$

as we saw in Note 35.

So the Hamiltonian, say $\mathcal{H} = p^2/(2m) + \mathcal{V}(x)$, becomes the operator

$$\mathcal{H} = -\frac{\hbar^2}{2m}\text{slope}_x^2 + \mathcal{V}(x)$$

where the operation associated with the $\mathcal{V}(x)$ is just to multiply the wavefunction by it.

So

$$\mathcal{H}\Psi(t, x) = \left(-\frac{\hbar^2}{2m}\left(-\frac{p^2}{\hbar^2}\right) + \mathcal{V}(x)\right)\Psi(t, x) = E\Psi(t, x)$$

But so also

$$i\hbar\,\text{slope}_t \Psi(t, x) = E\Psi(t, x)$$

So

$$i\hbar\,\text{slope}_t \Psi(t, x) = -\frac{\hbar^2}{2m}\,\text{slope}_x^2 \Psi(t, x) + \mathcal{V}(x)\Psi(t, x)$$

which is the *time-dependent Schrödinger equation*.

Since the time dependence is given by $e^{-iEt/\hbar}$ we can separate this out and define $\psi(x)$ by

$$\Psi(t, x) = e^{-iEt/\hbar}\psi(x)$$

30

This gives the *time-independent Schrödinger equation*

$$\frac{\hbar^2}{2m}\, \text{slope}_x^2 \psi(t,x) + (E - \mathcal{V}(x))\psi(x) = 0$$

In atomic physics (Note 30) we saw that $\psi(x)$ is not always $e^{ipx/\hbar}$. There $\psi(x)$, extended to polar coordinates in three dimensions, was given by a radial component times the spherical harmonics.

In general, $\psi(x)$ is found by solving the time-independent Schrödinger differential equation.

This usually requires more mathematical experience than we have developed in this course.

But sometimes we can find $\psi(x)$ more directly.

Let's try solving Schrödinger's equation for the quantum harmonic oscillator.

$$\mathcal{H} = \frac{kx^2}{2} + \frac{p^2}{2m}$$

For this discussion we replace k by $m\omega^2$, where $\omega = \sqrt{k/m}$, the factor that appears everywhere in the simple harmonic oscillator physics of Note 37.

$$\mathcal{H} = \frac{m\omega^2 x^2}{2} + \frac{p^2}{2m}$$

From Note 37 ω clearly is an angular frequency measured in radians/sec. It is not the ω of $e^{-i\omega t} = e^{-iEt/\hbar}$ in the wavefunction, but multiplying it by $\hbar$ gives it the dimensions of energy, which are the dimensions of $\mathcal{H}$. So let's work in energy units of $\hbar\omega = \hbar\sqrt{k/m}$.

$$\frac{\mathcal{H}}{\hbar\omega} = (x\sqrt{\frac{m\omega}{2\hbar}})^2 + (\frac{p}{\sqrt{2m\hbar\omega}})^2$$

I've written $\mathcal{H}$ with these square roots because it makes $\mathcal{H}$ the sum of two squares, which can be factored.

$$\frac{\mathcal{H}}{\hbar\omega} = (x\sqrt{\frac{m\omega}{2\hbar}} - \frac{ip}{\sqrt{2m\hbar\omega}})(x\sqrt{\frac{m\omega}{2\hbar}} + \frac{ip}{\sqrt{2m\hbar\omega}})$$

Woops. Not so fast! This would be OK if x and p were just numbers, but p is an operator, $p = -i\hbar\, \text{slope}_x$, and

$$[x, p] = i\hbar$$

Let's explore two operators which are going to turn out to be something like the lowering and raising operators of Note 30.

$$\begin{aligned} a &= x\sqrt{\frac{m\omega}{2\hbar}} + \frac{ip}{\sqrt{2m\hbar\omega}} \\ a^\dagger &= x\sqrt{\frac{m\omega}{2\hbar}} - \frac{ip}{\sqrt{2m\hbar\omega}} \end{aligned}$$

(Notice that they are conjugates of each other.)

Now

$$[a, a^\dagger] = 1$$

and

$$\frac{\mathcal{H}}{\hbar\omega} = a^\dagger a + \frac{1}{2}$$

(Check this!) So we weren't far off: the commutator just gives an extra $1/2$.

Furthermore

$$[\mathcal{H}, a] = -\hbar\omega a$$

$$[\mathcal{H}, a^\dagger] = \hbar\omega a^\dagger$$

In general, if we have an operator X whose commutator with the Hamiltonian gives a multiple of itself

$$[\mathcal{H}, X] = cX$$

then for some wavefunction ψ_n with energy E_n (i.e., $\mathcal{H}\psi_n = E_n\psi_n$)

$$\begin{aligned} \mathcal{H}X\psi_n &= (X\mathcal{H} + cX)\psi_n \\ &= XE_n\psi_n + Xc\psi_n \\ &= (E_n + c)X\psi_n \end{aligned}$$

so $X\psi_n$ is a new wavefunction with energy $E_n + c$.

Thus $a, a^\dagger$ create new wavefunctions, with energies

$$\begin{aligned} E_n - \hbar\omega &\quad \text{for} \quad a\psi_n \\ E_n + \hbar\omega &\quad \text{for} \quad a^\dagger\psi_n \end{aligned}$$

Furthermore, the energy is always $\geq \hbar\omega/2$ because the dot product of a vector with itself is never negative. (We'll write $|E>$ for a wavefunction of energy E in the following.)

$$\begin{aligned} 0 &\leq (a\,|\,E>, a\,|\,E>) \\ &= <E\,|\,a^\dagger a\,|\,E> \\ &= <E\,|\,\frac{\mathcal{H}}{\hbar\omega} - \frac{1}{2}\,|\,E> \\ &= <E\,|\,\frac{E}{\hbar\omega} - \frac{1}{2}\,|\,E> \end{aligned}$$

So

$$0 \leq \frac{E}{\hbar\omega} - \frac{1}{2}$$

Let's call the lowest energy state $E_0 = \hbar\omega/2$. Then the raising operator $a^\dagger$ generates, successively, $E_1 = \hbar\omega + \hbar\omega/2$, $E_2 = 2\hbar\omega + \hbar\omega/2$, .., $E_n = n\hbar\omega + \hbar\omega/2$, starting from the E_0 wavefunction.

Thus, now writing just $|n>$ for the wavefunction of energy $E_n = n\hbar\omega + \hbar\omega/2$

$$a^\dagger a\,|\,n> = (\frac{\mathcal{H}}{\hbar\omega} - \frac{1}{2})\,|\,n> = (\frac{E}{\hbar\omega} - \frac{1}{2}) = n\,|\,n>$$

So it must be that

$$a\,|\,n> = \sqrt{n}\,|\,n-1>$$

and

$$a^\dagger\,|\,n-1> = \sqrt{n}\,|\,n>$$

or

$$a^\dagger\,|\,n> = \sqrt{n+1}\,|\,n+1>$$

This justifies calling $a, a^\dagger$ lowering and raising operators respectively (or annihilation and creation operators, since they annihilate and create units of energy, respectively).

So without solving the differential equation we get the energy levels of the quantum harmonic oscillator.

In the pre-quantum limit, $\hbar \to 0$ so the levels get so closely spaced that they become a continuum of energies, just as we had in Note 37.

We can also get the wavefunctions themselves.

In the lowest ("ground") state $|\,0>$, $E_0 = \hbar\omega/2$ and we cannot go lower

$$
\begin{aligned}
0 &= a\,|\,0>= \left(x\sqrt{\frac{m\omega}{2\hbar}} + \frac{ip}{\sqrt{2m\hbar\omega}}\right)|\,0> \\
&= \left(x\sqrt{\frac{m\omega}{2\hbar}} + \sqrt{\frac{\hbar}{2m\omega}}\,\text{slope}_x\right)|\,0>
\end{aligned}
$$

This is a simple differential equation and we can solve it. It says

$$
\text{slope}_x\,|\,0>= -\frac{m\omega}{\hbar}x\,|\,0>
$$

so

$$
\text{antislope}_x \frac{\text{slope}_x\,|\,0>}{|\,0>} = \text{antislope}_x\left(-\frac{m\omega}{\hbar}x\right)
$$

from which, using the antislope form of the chain rule (Note 38)

$$
\text{antislope}_{|0>}\frac{\text{slope}_{|0>}\,|\,0>}{|\,0>} = \left(-\frac{m\omega}{\hbar}\right)\text{antislope}_x x
$$

or

$$
\text{antislope}_{|0>}\frac{1}{|\,0>} = \left(-\frac{m\omega}{2\hbar}\right)x^2
$$

From the rule for slopes of inverse functions (Note 30)

$$
\text{slope}_y \ln(y) = \frac{1}{\text{slope}_x e^x} = \frac{1}{e^x} = \frac{1}{y}
$$

where $y = e^x$, so

$$
\ln|\,0>= -\frac{m\omega}{2\hbar}x^2 + \text{const.}
$$

so finally

$$
|\,0>= \text{const } e^{-\frac{m\omega}{2\hbar}x^2}
$$

The constant is chosen to normalize $|\,0>$ so the area of $<0\,||\,0>$ from $-\infty$ to ∞ is 1

$$
\text{antislope } e^{-(x/a)^2}\,\big|_{-\infty}^{\infty} = a\sqrt{\pi}
$$

so

$$
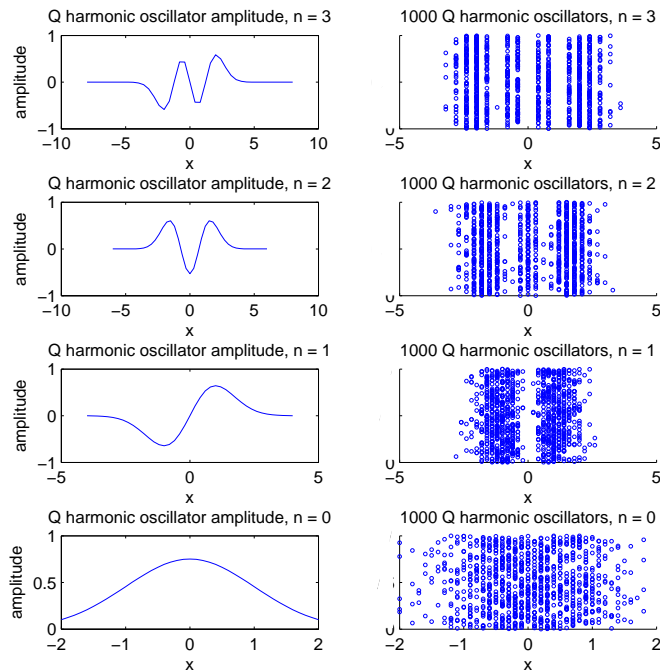|\,0>= \sqrt[4]{\frac{m\omega}{\pi\hbar}}\,e^{-\frac{m\omega}{2\hbar}x^2}
$$

From $\mid 0 >$ we can find $\mid 1 >, \mid 2 >, ..$: just apply $a^\dagger$ repeatedly

$$
\begin{aligned}
\mid 1 >= a^\dagger \mid 0 > &= (x\sqrt{\frac{m\omega}{2\hbar}} - \sqrt{\frac{\hbar}{2m\omega}}\text{slope}_x) \mid 0 > \\
&= \sqrt{2}\sqrt{\frac{m\omega}{\hbar}}x \mid 0 > \\
\mid 2 >= \frac{1}{\sqrt{2}}a^\dagger \mid 1 > &= \frac{1}{\sqrt{2}}((\sqrt{\frac{m\omega}{\hbar}}x)^2 - 1) \mid 0 > \\
\mid 3 >= \frac{1}{\sqrt{3}}a^\dagger \mid 2 > &= \frac{1}{\sqrt{3}}((\sqrt{\frac{m\omega}{\hbar}}x)^3 - \sqrt{\frac{m\omega}{\hbar}}x) \mid 0 >
\end{aligned}
$$

Here these are in pictures.

The relationship between quantum physics and pre-quantum physics is less straightforward than the relationship between timespace physics (special relativity) and pre-timespace physics.

Timespace physics can be worked out in its own right, and pre-timespace physics comes out as the special case in which $c \to \infty$.

Pre-quantum physics is the special case of quantum physics in which $\hbar \to 0$.

But quantum physics depends on pre-quantum physics in a way that timespace physics does not depend on pre-timespace physics.

This is because quantum physics has two components: the deterministic evolution of quantities in time, undisturbed by observation; and the disturbing act of observation or measurement.

The disturbance cannot be minimized because the "quantum of action" $\hbar$ which is the smallest possible action, is not zero.

(Those who use the phrase "quantum leap" to mean some kind of enormously transcendental

breakthrough should be careful. A quantum leap is correctly defined as the smallest possible leap.)

Measurement of a quantum system requires its interaction with an "apparatus", which Landau and Lifshitz [LL58, Sect.1] characterize as "a body of large enough mass". The philosophical issues of what happens during measurement are unresolved.

Remarkably, the mathematics of linear operators gives results whose agreement with empirical checking have not been faulted.

But measurement brings quantum phenomena into the ordinary realm of human beings, and thus is described by pre-quantum physics.

Thus the introduction of quantum physics in these Notes has not been from first principles, with pre-quantum physics derived at the end as a special case. Rather, we have shuttled back and forth between pre-quantum and quantum physics.

I hope to have minimized the shuttling and still provided good intuitions for every step.

40. Summary (These notes show the trees. Try to see the forest!)

Part I Discrete symmetries and molecules.

Notes 1–11. Symmetries of an equilateral triangle abstracted to groups. Invariant sets and subgroups. Traces and further matrix representations. Decomposing into irreducible representations and block-diagonalizing matrices.

Notes 12–17. Finding fundamental vibration modes of molecules from their symmetries: greenhouse gases CO_2 and H_2O.

Notes 16–18. Symmetries of the platonic solids: tetrahedron, octahedron/cube, dodecahedron/icosahedron.

Part II Infinite symmetries and crystals.

Notes 19–25. Translation symmetries and crystals in one and two dimensions: crystallography and waves.

Part III Continuous symmetries and the atom.

Notes 26–29. Rotational symmetry in two and three dimensions. Commutator algebra and representations of the spherical group.

Notes 30, 31. Spherical harmonics and atomic physics.

Part IV Abstract symmetries and lots of physics

Notes 32–34. From SU(2) (the atom) to SU(3) (the quarks). Isospin and hypercharge.

Notes 35–39. Symmetry and conservation laws. Complementary quantities, energy, Lagrangian, principle of stationary action, Noether's theorem, Hamiltonian, Schrödinger's equation and the quantum harmonic oscillator.

II. The Excursions
You've seen lots of ideas. Now *do* something with them!

1. Show that $(X^2)^\dagger = (X^\dagger)^2$. Hence argue that $(e^{iX})^\dagger = e^{-iX^\dagger}$. (Note 32.)

2. Why are reflections and inversions excluded from the special unitary groups SU(2), SU(3), ..? How can they be put in?

3. Show that $[T_a, T_b] = if_{abc}T_c$ for the adjoint representation of Note 32 because $f_{bcd}f_{ade} + f_{abd}f_{cde} + f_{cad}f_{bde} = 0$ because $[X_a, [X_b, X_c]] + [X_b, [X_c, X_a]] + [X_c, [X_a, X_b]] = 0$ (Note 28) and $[X_a, [X_b, X_c]] = if_{bcd}[X_a, X_d] = -f_{bcd}f_{ade}X_e$

4. Show that the eight half-Gell-Mann matrices are hermitian (so the group they generate is unitary), traceless (the group is special) and independent.

5. $I_z + U_z + V_z = 0$ means that $I_z v$, $U_z v$ and $V_z v$ are coplanar for any vector $v = (x, y, z)$. For $v = (1, 1, 1)$ find the transformation that maps $I_z v$, $U_z v$ and $V_z v$ respectively onto $(0, -1, 0)$, $(a, b, 0)$ and $(c, d, 0)$, and find a, b, c and d with $a^2 + b^2 = 1 = c^2 + d^2$. Draw the resulting points in two dimensions, Are there other (x, y, z) that can be transformed to the same picture? Find an (x, y, z) that cannot.

6. In Note 33, relate the SU(3) commutators $[U_x, U_y]$ and $[V_x, V_y]$ to $(I_z + 2U_z)/\sqrt{3} \pm Y_\pm$

7. Calculate the adjoint representation matrices for SU(3), especially for I_z and Y, directly from the commutators. Show that, when I_z and Y are diagonalized, their elements are those shown in Note 33.

8. Why does the adjoint representation of SU(3) not give rise to a second 8D representation by taking 2-number (complex) conjugates?

9. If the directions of the grid edges in Note 33 did not matter, how many different grids are there with threefold rotational symmetry about the origin?

10. Find a triangular "decuplet" (10-point) representation of SU(3), $xu = 30$, in the grid for the adjoint representation in Note 33.

11. Show the smallest four representations for SU(3) on the grid of the adjoint representation, $xu = 11$, in the manner used to show 10, 02, 21 and 13 in Note 33.

12. a) Confirm the coordinates of $(x/2, (x/2 + u)/\sqrt{3})$ for the rightmost top point of the SU(3) representations illustrated in Note 33.
b) Conversely show that if we know the coordinates, say (z, y), of the rightmost top point, then $x = 2z$ and $u = \sqrt{3}(y - 2x)$.

13. Given the I_z and Q values for the baryon and meson octets in Note 34, confirm their hypercharges Y.

14. Check that the dimensions $(T, L, M$—see Week 7a Note 8$)$ are the same for $\mathcal{T} = mv^2/2$ and $\mathcal{V} = mgh$ in Note 36. This is the dimension of energy.

15. a) If the hill in Note 36 is 5 meters high, how fast is the ball moving at the bottom, assuming no energy has been lost through friction (heat) as it rolls from top to bottom?
b) Can you think of a way of measuring v to check the formula?

16. In Note 36 we said two apparently conflicting things about energy. For the ball rolling down the hill we said $mv^2 = mgh$, i.e., apparently, $\mathcal{T} = \mathcal{V}$. Then we said $E = \mathcal{T} + \mathcal{V}$. Why are these not contradictory?

17. Note 36 points out that Note 7 of Week 5 was sloppy in considering only time and not distance. It was also sloppy, as a result, in not specifying that all paths must start and end at the same time or else the interference could not happen.
a) Show that velocity $v(j) = 2\sqrt{L^2 + (jd)^2}/(t_2 - t_2)$ (where L is the length given in Week 5) varies for different paths j and that the calculation of Week 5 assumes that all the dependence of $v(j)$ on j is taken on by ω in $v = \omega/k$.
b) Given the definition in Note 36 of $\int_{t_1}^{t_2} \mathcal{L}dt$ as the area from t_1 to t_2 under the curve described by $\mathcal{L}$, show that, in Week 5 Note 7

$$\int_{t_1}^{t_2} \omega(j)dt = (t_1 - t_2)\omega(j) = 2k\sqrt{L^2 + (jd)^2}$$

(L is the Week 5 length again) for path j. (In Week 5, ω does not change with t.)

c) Setting $\text{slope}_j \int_{t_1}^{t_2} \omega(j)dt$ to zero, show that $\int_{t_1}^{t_2} \omega(j)dt$ is stationary at $j = 0$.

d) Look up Newton's laws and show that the above sort of thinking leads to Newton's first law.

e) Show that conservation of momentum gives Newton's third law.

18. Partial integration can be tricky. Where does te following proof that $0 = 1$ go wrong?

$$\text{antislope } \frac{1}{x}1 = \frac{1}{x}x + \text{antislope } \frac{x}{x^2} = 1 + \text{antislope } \frac{1}{x}$$

Hint: go back to the derivation of partial integration in Note 37.

19. **Billiards.** Since we have now derived Newton's three laws, in principle we know all of "classical" (pre-relativity, pre-quantum) physics. To experience it we need practice. A paradigmatic exercise is the collision of billiard balls. We will use conservation of momentum and we will assume the collision conserves energy of motion. Such a collision is called *elastic*. Conversely an inelastic collision absorbs some of the energy of motion as heat or excitation of atomic levels or increasing the mass of the colliding objects, and does not give it all back as energy of motion after the collision. We saw relativistic examples of both, in a single dimension of space, in Week 7a Notes 11–15, having *assumed* conservation of energentum. Now we have *proved* conservation of momentum and of energy separately for pre-relativistic speeds. And we can move on to higher dimensions of space.

a) Let's start with an actual billiards example. In billiards, all the balls are stationary when a player cues one, which then usually collides with (at least) one other (maybe). Suppose the cue ball has momentum $\vec{p_1}$ while the ball it hits has momentum $\vec{0}$. We want to find the outgoing momenta, $\vec{p_1'}$ and $\vec{p_2'}$.

Conservation of momentum says

$$\vec{p_1'} + \vec{p_2'} = \vec{p_1}$$

This gives a triangle of the three vectors (draw it!).

Conservation of energy says (billiard balls all ideally have the same mass, m, so we can factor the $2m$ out of the energy $= p^2/(2m)$.)

$$p_1'^2 + p_2'^2 = p_1^2$$

This makes the triangle pythagorean, and we can see that the outgoing momenta are orthogonal (i.e., make a right angle with each other), a property a good billiards player may try to exploit.

So far we have three equations (the vector equation for momenta is two of them) in four unknowns (the two components each of $\vec{p_1'}$ and $\vec{p_2'}$).

The fourth equation we need is given by the direction of the line joining the centres of the two balls at the moment of collision, $\vec{c_{12}} = \vec{c_1} - \vec{c_2}$. This must also be the direction, positive or negative, of the changes in momenta, $\vec{p_1} - \vec{p_1'}$ for ball 1, and $\vec{p_2'}$ itself for ball 2. (This is just conservation of momentum with the components resolved in the direction of $\vec{c_{12}}$ and perpendicular to it, plus the fact due to absence of friction—or else the collision would not be elastic—that the momenta perpendicular to $\vec{c_{12}}$ cannot change.) In the two dimensions of the billiard table we can write directions as slopes:

$$\frac{\vec{p_{2y}'}}{\vec{p_{2x}'}} = \frac{\vec{c_{12y}}}{\vec{c_{12x}}} = \sigma$$

where we define σ as shorthand for the second fraction.

Show that these four equations now give

$$\vec{p_1} = \frac{p_{1y}' - \sigma p_{1x}'}{1 + \sigma^2}(-\sigma, 1)$$

37

$$\vec{p_2} = \frac{\sigma p'_{1y} + p'_{1x}}{1 + \sigma^2}(1, \sigma)$$

What must we do if $\vec{c_{12x}} = 0$?

Solve the example $\vec{p'_1} = (4, 3)$ and $\sigma = 0$.

Illustrate time-reveral symmetry by showing that the physics of elastic collisions works backwards: reverse the outgoing momenta you just found and use them as incoming momenta.

b) This derivation is easy to generalize to the case in which the target ball is not stationary: if it initially has momentum $\vec{p'_2}$ just subtract this from $\vec{p'_1}$ to get a new $\vec{p'_1}$ which we can plug into the above equations, and then add $\vec{p'_2}$ back to the results above. In this way you have temporarily pretended that you are moving along with ball 2 so that it is stationary from your moving point of view. In doing so you invoked Galilean relativity (Week 3, Note 11), a meta-principle which transcends Newton's laws (and was used by Einstein to establish that lightspeed is the same for all observers). Note that the new outgoing momenta, $\vec{p_1} - \vec{p'_2}$ and $\vec{p_2} - \vec{p'_2}$, are no longer orthogonal in general.

Solve the example $\vec{p'_1} = (3, 4), \vec{p'_2} = (-1, 1)$ and $\sigma = 0$.

Try finding the general result using the momentum, energy and direction equations directly given two nonzero initial momenta, $\vec{p'_1}$ and $\vec{p'_2}$.

c) It would be nice to find a simpler derivation and even nicer to find one which is not limited to two dimensions. The physicist's trick is to go to *centre-of-mass* (CoM) coordinates by subtracting $\vec{p'_{12}} = (\vec{p'_1} + \vec{p'_2})/2$ from both $\vec{p'_1}$ and $\vec{p'_2}$ beforehand to get new $\vec{p'_1}$ and $\vec{p'_2}$ (and then to go back by adding $\vec{p'_{12}}$ to both $\vec{p_1}$ and $\vec{p_2}$ afterwards).

The advantage is that in CoM

$$\vec{p'_1} + \vec{p'_2} = 0 = \vec{p_1} + \vec{p_2}$$

so that $\vec{p_2} = -\vec{p_1}$ and $\vec{p'_2} = -\vec{p'_1}$ and so from conservation of energy

$$p_1^2 + p_2^2 = p_1'^2 + p_2'^2$$

$p_1^2 = p_1'^2$ and $p_2^2 = p_2'^2$ so both incoming and outgoing momenta for each ball lie on a circle centered at that ball at the moment of collision. As in (a) the effect of conservation of momentum and absence of friction especially perpendicular to $\vec{c_{12}}$ now mean that $\vec{p_1}$ is the negative reflection of $\vec{p'_1}$ in $\vec{c_{12}}$ (negative because $\vec{p'_1}$ is incoming but $\vec{p_1}$ is outgoing) and similarly for $\vec{p_2}$ and $\vec{p'_2}$.

Interval algebra (Week 7c Note 8) gives us the answer immediately

$$\vec{p_1} = -\frac{\vec{c_{12}}\vec{p'_1}\vec{c_{12}}}{c_{12}^2}$$

$$\vec{p_2} = -\frac{\vec{c_{12}}\vec{p'_2}\vec{c_{12}}}{c_{12}^2}$$

Alternatively we can use dot products to give the component of $\vec{p'_1}$ parallel to $\vec{c_{12}}$, and hence its negative which is the component of $\vec{p_1}$ parallel to $\vec{c_{12}}$—so $\vec{p_1}$ is $\vec{p'_1}$ minus twice this component— and similarly for $\vec{p_2}$

$$\vec{p_1} = \vec{p'_1} - 2\vec{c_{12}}\frac{\vec{p'_1} \cdot \vec{c_{12}}}{c_{12}^2}$$

$$\vec{p_2} = \vec{p'_2} - 2\vec{c_{12}}\frac{\vec{p'_2} \cdot \vec{c_{12}}}{c_{12}^2}$$

Note that this works in any number of dimensions, not just two. (What happens to the interval algebra above in > 2 dimensions?)

Solve the example, already in CoM, $\vec{p_1} = (2, 3/2) = -\vec{p_2}$ and $\delta = 0$.

d) If you play billiards as often as I do (maybe twice in my life, when I happened to be near a billiard table with friends who like to play—which will explain any clumsiness in my use of billiards vocabulary) your main problem will be hitting the target ball. *You* do not need to determine what happens after you succeed: nature determines that for you. So let's see what it takes to decide whether incoming balls with momenta $\vec{p_1}$ and $\vec{p_2}$ actually will collide, and what $\vec{c_{12}}$ will be when they do. (But I'm going to take half of this and call it $\vec{r}$ because we will be changing the meanings of $\vec{c_1}$ and $\vec{c_2}$.)

If the balls have zero radius, finding out if and when they collide is easy. Given they start with their centres at $\vec{c_1}$ and $\vec{c_2}$ (this is a different meaning for these symbols than in (a) above) then their positions at time t will be $\vec{c_1} + \vec{v_1}t$ and $\vec{c_2} + \vec{v_2}t$ and these must coincide for a collision. (We must now switch from momenta $\vec{p_i}$ to velocities $\vec{v_i} = \vec{p_i}/m$ in classical physics, but we keep all the masses, m, equal as before.) $t(\vec{v_2} - \vec{v_1}) = \vec{c_1} - \vec{c_2}$ means

$$\frac{c_{1x} - c_{2x}}{v_{2x} - v_{1x}} = \frac{c_{1y} - c_{2y}}{v_{2y} - v_{1y}} = t$$

where the first equality must be true for the second to give t.

Giving the balls finite radii means we must draw "railway tracks", giving the outside positions of both balls, and we must then look at all the possible positions those balls can be in once they reach the diamond-shaped intersection where they might collide. Try putting these words into drawings and explore the possible directions $\vec{r}$ can take at the point of impact.

e) This gets complicated, is limited so far to two dimensions, and needs special treatment if either component of the velocity differences equals zero. So we remember that we saved work before by switching to CoM, and we do it again. This will make the collision point stationary at the origin.

We also get rid of the "railway tracks" by tracking only the centre of each ball but looking for where it hits a sphere of radius r (the radius of each of the balls) centered at the origin. Check that this contains all the information that the railway tracks did.

Since the two balls travel with exactly opposite velocities (if their masses are equal) in CoM, we need only find if and where and when one ball hits the sphere. This will happen for some vector $\vec{r}$ of length r if

$$\vec{c} + \vec{v}t = \vec{r}$$

from which

$$r^2 = (\vec{c} + \vec{v}t) \cdot (\vec{c} + \vec{v}t) = c^2 + 2\vec{c} \cdot \vec{v}t + v^2 t^2$$

so

$$t = \frac{-\vec{c} \cdot \vec{v} \pm \sqrt{(\vec{c} \cdot \vec{v})^2 - (c^2 - r^2)}}{v^2}$$

If the discriminant $(\vec{c} \cdot \vec{v})^2 - (c^2 - r^2) < 0$ then the balls miss. Otherwise, take the smaller solution because the larger one corresponds to the balls partially passing through each other until their centres are again separated by r.

The vector connecting the centres of the balls at the moment of collision is $2\vec{r}$. We need it for (a), (b) and (c) above.

The values for $\vec{c_i}$ and $\vec{v_i}$ can be taken to be their values at the moment of their previous collisions. (Why does it not matter that both balls probably did not have their previous collisions at the same time?)

Solve the example, already in CoM, $\vec{v_1} = (2, 3/2) = -\vec{v_2}$, $\vec{c_1} = -(21, 15)$ and $r = 1$. What are the two times and the two results for $\vec{r}$?

f) Generalize the discussions of both (c) and (e) above to balls with different masses and

different radii.

g) Because billiard balls move exceedingly slower than light (and so do gas molecules) we can separate energy from momentum to very good approximation. "But the observer is stationary with respect to the billiard table (or the gas container) and so we can safely separate the components of energentum." Why does this last statement lead us astray in making the relativistic version of the above calculations?

20. Show that symmetry of the Lagrangian under timespace translation gives conservation of energentum (used in Week 7a).

21. Potential energy is an intriguing concept. You must do work to acquire it—climbing the hill or compressing the spring—and this requires energy Fx to work for a distance x against force F.
    But potential energy is relative. Where it equals zero is an arbitrary choice—at the bottom of the hill? at infinitely far from the centre of force in atoms and planetary systems?
    So if you could change the local zero of your own potential energy, where would you set it to make yourself bulletproof?

22. In Note 37 why does the mass move clockwise on the x-$\dot{x}$ ellipse?

23. Modify the MATLAB program SHOquarter.m (in MATLABpak) to find the cubic spline of Note 37
    a) from $t = 0$ to $t = t_2$
    b) from $t = 0$ to $t = 2t_2$
    c) from $t = 0$ to $t = 4t_2$
    What is the action in each case?

24. Use antislope$_f$ slope$_f g(f) = g(f) = g(f(x)) = $ antislope$_x g(f(x))$ and the chain rule for slopes to prove the antislope version of the chain rule.

25. Draw arrows showing slope$_x \mathcal{V}(x)$, i.e., without the $-$ sign, on the last figure of Note 37.

26. For the simple harmonic oscillator in Note 37 show that $m\ddot{x} + kx = 0$
    a) from the Euler-Lagrange equation
    b) using the function for $x(t)$ that gives the motion of the oscillator following the elliptic path in x-$\dot{x}$ space.

27. In deriving the Euler-Lagrange equation from the principle of stationary action in Note 37 we implicitly assumed that both slopes and antislopes are linear operators, viz

$$\text{slope}_x(af(x) + bg(x)) = a\,\text{slope}_x f(x) + b\,\text{slope}_x g(x)$$

and similarly for antislope.
From the definition slope$_x f(x) \approx (f(x + \Delta x) - f(x))/\Delta x$ show that these linearities hold.

28. Read Feynman's [FLS64, Ch.19]. Who was the "student of that same teacher, Bader, I mentioned at the beginning of this lecture" who discovered in 1942 that quantum mechanics has a principle of least action? Relate this to the description of photons in [Fey85].

29. What interval algebra expression contains both $xp_x + yp_y$ and $xp_y - yp_x$? (Note 38 is the context for this question.)

30. On the principle of least action look up Count Pierre Louis Maupertuis 1698–1758, Leonhard Euler 1707–83, Count Joseph Louis Lagrange 1736–1813 and Willam Rowan Hamilton 1805–65. Why did Lagrange in particular not pursue his insights and arrive at quantum mechanics 135 years ahead of his time?

31. What is the Lagrangian for the sun-planet system? For the Bohr atom?

32. Work through the derivation of Noether's theorem in [Bae02].
Look up Emmy Noether (1882–1935). What problem with energy conservation in general relativity motivated her work on symmetry?

33. Look up the "Hermite polynomials" and compare them to the quantum harmonic oscillator wavefunctions of Note 39.

34. Show that the operator $\dot{f} = \text{slope}_t f + (i/\hbar)[\mathcal{H}, f]$, where f is any operator representing a physical quantity and $\mathcal{H}$ is the Hamiltonian, results from redefining $\text{Slope}_t f$ as follows [LL58, Sect.9].
Since $\text{Slope}_t f$ in pre-quantum physics means the rate of change of f in time, but since physical values in quantum physics do not generally have successive values in time, we must redefine

$$
\begin{aligned}
\text{Slope}_t f &= \text{Slope}_t(\text{mean value of } f) \\
&= \text{Slope}_t <\mid f \mid> \\
&= (\text{slope}_t <\mid)f \mid> + <\mid (\text{slope}_t f) \mid> + <\mid f(\text{slope}_t \mid>)
\end{aligned}
$$

Then, remembering that $<\mid$ involves conjugates of $\mid>$, along with any operations applied to these, we can use $\mathcal{H} \mid>= i\hbar\, \text{slope}_t \mid>$ and $<\mid \mathcal{H} = -i\hbar\, \text{slope}_t <\mid$.
Show that the result follows if we define $\dot{f}$ as the quantity whose mean value equals $\text{Slope}_t f$:
$<\mid \dot{f} \mid>= \text{Slope}_t f$.
Tidy up this derivation by removing "$\text{Slope}_t f$" from it altogether.
What is the result if f does not explicitly depend on t?

35. Show that the relativistic version of the Schrödinger equation for a free particle is as given in Note 16 of Week 7a.
This is called the Dirac equation for a free particle and is discussed in [Dir58, Sect.67].
Since $E^2 - p^2 = m^2$ the Hamiltonian will be given by $\mathcal{H} = \sqrt{p^2 + m^2}$ and so
$(\mathcal{H} - \sqrt{p^2 + m^2})\psi = 0$.
The square root makes it difficult to apply the operator $p = -i\hbar\, \text{slope}_x$ in 1D and
$\vec{p} = -i\hbar(\text{slope}_x, \text{slope}_y, \text{slope}_z)$ in 3D.
We can multiply both sides by $\mathcal{H} + \sqrt{p^2 + m^2}$ to get $(\mathcal{H}^2 - (p^2 + m^2))\psi = 0$ and the solutions will include the solutions of the original wave equation. But [Dir58, Sect.27] shows that the time evolution of the wavefunction, $\mathcal{H}\psi = E\psi$, must be a linear operator. So the quadratic appearance of $\mathcal{H} = i\hbar\, \text{slope}_t$ is just as problematical as the square root over p^2 was before.
So Dirac looked for a linear equation of the form $(\mathcal{H} - (\vec{\alpha} \cdot \vec{p} + \alpha_4 m))\psi = 0$ (where $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ in 3D). The α_i must not depend on position (x, y, z) for a free particle, to conserve momentum, and so $\vec{\alpha}$ commutes with both position and momentum.
They do not commute with each other, so if we try to get $(\mathcal{H}^2 - (p^2 + m^2))\psi = 0$ by multiplying by $(\mathcal{H} + \vec{\alpha} \cdot \vec{p} + \alpha_4 m)$:

$$(\mathcal{H} + (\vec{\alpha} \cdot \vec{p} + \alpha_4 m))(\mathcal{H} - (\vec{\alpha} \cdot \vec{p} + \alpha_4 m))\psi = 0$$

we need

$$\alpha_j \alpha_k = -\alpha_k \alpha_j \quad j, k = 1, 2, 3, 4$$

and

$$\alpha_j^2 = 1 \quad j = 1, 2, 3, 4$$

a) Show this.
These anticommutative properties remind us of the Pauli matrices (Week 6 Note 5)

$$
\sigma_1 = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \quad
\sigma_2 = \begin{pmatrix} & -i \\ i & \end{pmatrix}, \quad
\sigma_3 = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}
$$

$$\sigma_j \sigma_k = -\sigma_k \sigma_j \quad j, k = 1, 2, 3$$
$$\sigma_j^2 = 1 \quad j = 1, 2, 3$$

but there are only three of these.

b) Show that there is no set of four 2×2 matrices that anticommute the way the αs should.

c) Show that the 4×4 matrices in Week 7a Note 14 (with $\beta = \alpha_4$) anticommute as Dirac desired.

d) The first chapter of [Dir58] is elegant but following it requires prior experience of quantum mechanics. The discussions of Week 7a and here should help. Try reading it now.

36. Look up the discussion of the passage to the limiting case of classical mechanics in [LL58, Sect.6], [Dir58, Sect.12] and [FLS64, Ch.19].
Note the use of the pre-quantum action S in the wavefunction, but divided by $\hbar$, meaning that the phase of the wavefunction oscillates so rapidly that the motion can be taken to be that of a particle moving in a single line (straight if it is a free particle). Note the comparison with geometrical (ray) optics as the limiting case of wave optics.

37. Any part of the lecture that needs working through.

References

[Bae02] John Baez. Noether's theorem in a nutshell. math.ucr.edu/home/baez/noether.html Accessed 09/1/21, March 12 2002.

[Dir58] P. A. M. Dirac. *The Principles of Quantum Mechanics.* Oxford University Press, Oxford, 1958. 4th ed.

[Fey85] R. P. Feynman. *QED: The Strange Theory of Light and Matter.* Princeton University Press, Princeton, N.J., 1985. Alix G. Mautner Memorial Lecture, UCLA, 1983.

[FLS64] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics, Volume II.* Addison-Wesley, 1964.

[LL58] L. D. Landau and E. M. Lifshitz. *Quantum Mechanics: Non-relativistic Theory.* Pergamon Press Ltd, Oxford, 1958. Vol 3 of Course of Theoretical Physics; Translated by J. B. Sykes and J. S. Bell.